CHAPTER 1

Introduction to Statistics and Levels of Measurement

How to figure things out.

OBJECTIVES

By the end of this chapter students will be able to:

- State the question that statistics is always trying to answer.
- Define the empirical method.
- Compare quantitative and qualitative variables.
- Differentiate a population from a sample and a statistic from a parameter, giving an example of each.
- Explain the difference between an independent and a dependent variable, citing examples of each.
- Identify continuous and categorical variables accurately.
- Distinguish the four levels of measurement, and describe each.
- Apply several beginning-level statistical techniques to further develop understanding of the concepts discussed in this chapter.

KEY TERMS

Categorical variable
A variable that has a finite number of classification groups or categories, which are usually qualitative in nature.

Continuous variable
A variable that has an infinite number of potential values, with the value being measured falling somewhere on a continuum containing in-between values.

Dependent variable
The outcome variable or final result.

Empirical method
A way of gathering information through systematic observation and experimentation.

Estimate
A preliminary approximation.

Independent variable
A variable measured or controlled by the experimenter; the variable that is thought to affect the outcome.

Interval data
Data whose categories are exhaustive, exclusive, and rank-ordered, with equally spaced intervals.

Nominal data
Data that indicates a difference only, with categories that are exhaustive and exclusive but not rank-ordered.

Ordinal data
Data whose categories are exhaustive, exclusive, and rank-ordered.

Parameter
Descriptive result for the whole group.

Population
The whole group.

Probability
The likelihood that an outcome will occur.

Qualitative measure
A measure that describes or characterizes an attribute.

Quantitative measure
A measure that reflects a numeric amount.

Ratio data
Data whose categories are exhaustive, exclusive, and rank-ordered with equally spaced intervals and a point at which the variable does not exist.

Sample
A group selected from the population.

Statistic
An estimate derived from a sample.

Variable
The changing characteristic being measured.

INTRODUCTION

So here you are. You've worked hard, you are in nursing school, and you are ready to begin. But wait! Why do you have to take statistics? Why do you need to understand all those numbers and equations when you are a nurse and want to help people?

Most nursing students experience a mild sense of panic when they discover they have to take statistics—or any other kind of math, for that matter. That reaction is common. Here is a calming thought to remember: You already practice statistics, but you just may not know it.

Statistics boils down to doing two things:

- Looking at data
- Applying tests to find out either (1) that what you observe is what you expected or (2) that your observation differs enough from what you expected that you need to change your expectations

You might be convinced that you don't use statistics in your life, so let me give you an example. New York State, where I live, has four seasons. The summer is usually June, July, and August. Fall is September, October, and November. Winter is December, January, and February. And that leaves March, April, and May for the spring. If you walk outside in July and find it to be 80° and humid, you would draw an unspoken conclusion that what you just observed is what you were expecting, and you would put on your sunglasses. However, what if you walk outside in January and find it to be 80° and humid? You would probably be startled, take off your overcoat and boots, and read about global warming. The difference between the weather you expect in winter and what you actually encounter is so different that you might need to change your expectations. You are already practicing statistics without knowing it!

Of course, that day in January might just be a fluke occurrence (a random event), and the temperature could be below freezing again the next day. That is why we need to use the **empirical method**, otherwise known as systematic observation and experimentation. The empirical method allows you to determine whether the temperature observed is *consistently* different from what you expect. To use the empirical method, you need to check the daily temperature on more than one day. So you might decide to monitor the daily temperature for a whole month of winter to see whether readings are consistently different from what you expect in the winter months. In this scenario, you would be using the empirical method to practice statistics (see **Figure 1-1**).

FIGURE 1-1 **2017 Winter Outlook.**

Courtesy of Farmers' Almanac.

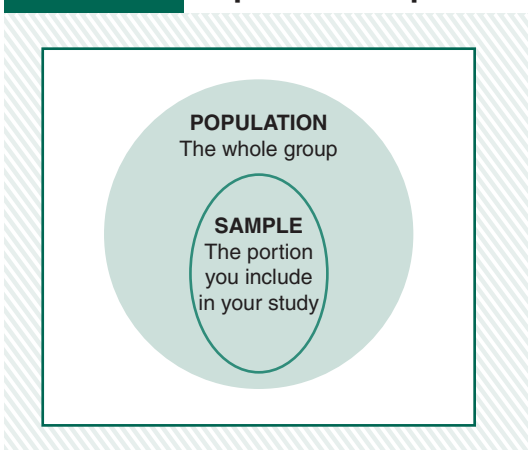
POPULATION VERSUS SAMPLE

To answer questions in research, we need to set up a study of the concepts we're interested in and define multiple **variables**, that is, the changing characteristics being measured. In our example, the temperature is a variable, a measured characteristic. Each variable has an associated **probability** for each of its possible outcomes, that is, how likely it is the outcome will occur. For example, how likely is it that the temperature will be below freezing as opposed to being in the eighties in winter? In your study, you recorded the daily temperature for a winter month, and those readings make up a **sample** of all the daily temperatures in the months of winter. The manner in which you collect your sample is dependent on the purpose of your study (**Figure 1-2**).

A sample is always a subset of a **population**, or an overall group (sometimes referred to as the reference population). In this case, our population includes all the daily temperatures in the winter months, and the subset, or sample, is all the daily temperatures recorded during your month of data collection. If you calculate the average temperature based on this sample data, you create what is called a **statistic**, which is an **estimate** generated from a sample.

A measured characteristic of a population is called a **parameter**. In our example, if you measured the daily temperature for December, January, and February and then calculated the average temperature, you would be determining a parameter. A really good way to remember the relationships among these four terms is with the following analogy: Statistic is to sample as parameter is to population.

FIGURE 1-2 **Population vs. Sample.**



QUANTITATIVE VERSUS QUALITATIVE

While you are collecting the weather data, you may realize that the data can be recorded in several ways. You could write down the actual temperature on that day, which would be a **quantitative measurement**, or you could describe the day as "warm" or "cold," which would be a **qualitative measurement**. A numeric amount or measure is associated with quantitative measurement (such as 80°F), and qualitative measures describe or characterize things (such as, "So darn cold I can't feel my toes").

Be careful with this difference: You can easily get confused. Qualitative variables do not contain quantity information, even if numbers are assigned. The assigned numbers have no quantitative information, rank, or distance. For example, a survey question asks, "What color scrubs are you wearing?" and lists choices numbered 1 to 3. Even if you selected choice 2, neon orange, you do not necessarily have any more scrubs than someone who chooses 1, lime green (although both respondents may want to purchase new scrubs). Even though these qualitative variables have numbers assigned to them, the numbers simply help with coding. The variables are still qualitative.

Independent versus Dependent Variables

Being as inquisitive as you are, you have probably asked yourself a number of times about a relationship you observe in your patients. For example, you notice that many supportive family members visit Sally Smith after her hip replacement recovery and that she is discharged 3 days after her surgery. Joanne Jones, on the other hand, has no visitors during her hip replacement recovery and is not discharged until day 6. As an observant nurse researcher, you have been wondering how variable *x* (the **independent variable**, which is measured or controlled by the experimenter) affects variable *y* (the **dependent variable**, or outcome variable) (**Figure 1-3**). You wonder, does having family support (the independent variable) affect the duration of a hospital stay (the dependent, or outcome, variable)?

To answer this question, you create a study. Obviously, other factors might be involved as well, but in your experiment, you are interested in how family support, the independent variable, affects hospital stay, the dependent variable. If you are correct, then the duration of the hospital stay *depends* on family support. The independent variable can be a suspected causative agent, and the dependent variable is the measured outcome or effect (**Figure 1-4**).

FIGURE 1-3 **Relationship of Independent and Dependent Variables.**

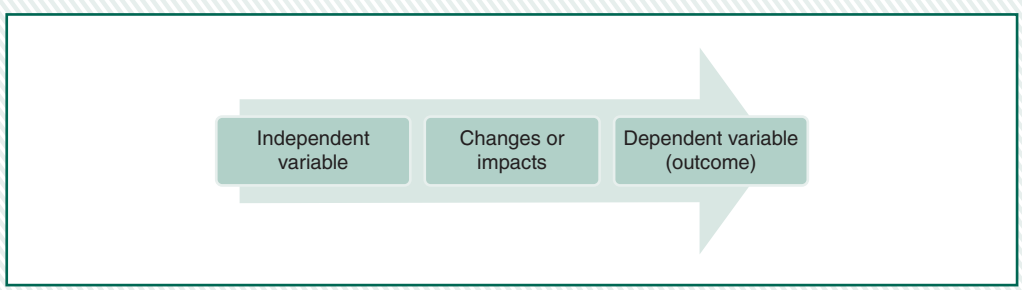
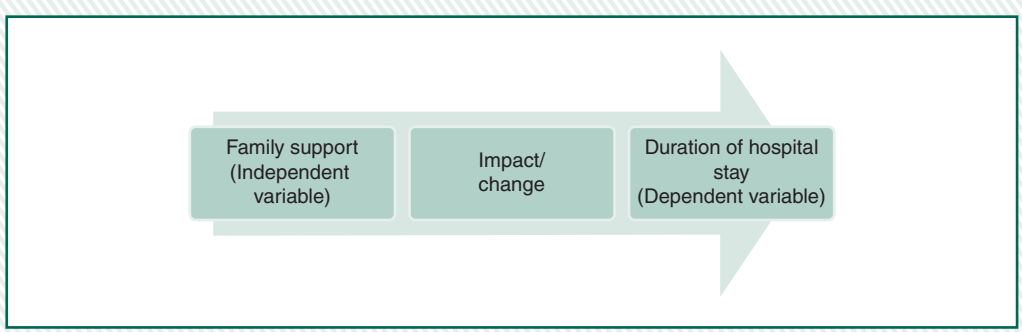


FIGURE 1-4 **Does Family Support Affect the Duration of a Subject's Hospital Stay?**



Note: Additional criteria must be met to say that a variable is causative, so I refer here only to the "suspected" causative agent.

Continuous versus Categorical Variables

Some data have an infinite number of potential values, and the value you measure falls somewhere on a continuum containing in-between values. These values are called **continuous variables**. As a nurse, when you measure your patient's temperature, you are measuring a continuous variable. The reading could be 98° or 98.6° or 98.66666°. The infinite possibilities are all quantitative in nature. Actually, the only limit to the measurement is the accuracy of the measuring device. For example, if you have a thermometer that measures only in whole degrees, you will not have as much information as you would using a thermometer that measures to the one-thousandth of a degree.

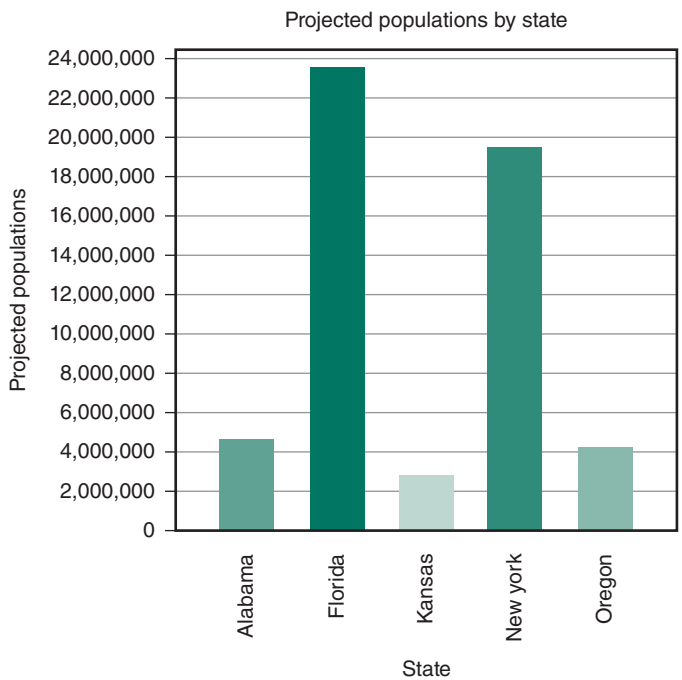
Continuous variables can be contrasted with **categorical variables**, sometimes called discrete variables, which have a finite number of classification groups, or categories, that are usually qualitative in nature. For example, as part of your research you may need to collect information about your patients' racial background. The choices available are African American, Native American, Caucasian, Asian, Latino, mixed race, and other. Race is an example of a categorical variable, a measurement that is restricted to a specific value and does not have any fractional or in-between values. When you read a study, the demographic information about the sample involved usually contains quite a few categorical variables including marital status, gender, race, geographic region, educational level, language spoken, smoking status, and so on.

Let's look at an example where we can see both types of variables in a study. If you were reading a public health study examining statewide variation in population estimates you might have the information in **Figure 1-5** available. Your sample was collected and reported about five states, so the state becomes one of the demographic variables you will want to report. Note that "state" is a categorical/qualitative variable: It just tells you the location of the sample subject and does not include any quantitative information. You also record the state population, which is a continuous/quantitative variable where the value can fall anywhere within the range of population values.

Levels of Measurement

Let's say that your interest in the relationship between family support (the independent variable) and duration of stay (the dependent variable) is extensive enough that you apply for a program at your hospital that includes a small research fellowship. You win the fellowship and proceed to collect data about each patient admitted to your orthopedic unit for hip replacement over a 3-month period. The study protocol calls for you to complete the usual admission forms and then for patients to complete a short survey about perceived family support. After your institutional review board approves your

FIGURE 1-5 **Projected Populations by State.**



Reproduced from CDC. (n.d). Population Projections, United States, 2004–2030, by state, age and sex. CDC WONDER Online Database, September 2005. Accessed at http://wonder.cdc.gov/population-projections.html on Apr 4, 2017.

study, you begin. The level of measurement of your data determines what type of analysis you are able to perform in your study, so let's look at the different types and what makes each level unique.

Your first survey question asks the patient's gender (male, female, other). The data you gather for this question is an example of **nominal data**; it simply indicates a difference between the three answers. One is neither greater than nor less than the other, and they are not in any particular order. Also, the categories are exclusive and exhaustive; that is, the patient cannot answer "both" or "neither." Asking about

FROM THE STATISTICIAN *Brendan Heavey*

What Is a Statistic?

As a student of statistics, you will run into questions regarding parameters and statistics all the time. Determining the difference between the two can be difficult. To get a concrete idea of the difference, let's look at an example. According to the Bureau of Labor Statistics, registered nurses constitute the largest healthcare occupation, with 2.7 million jobs nationwide. Because this text is primarily designed for nursing students, let's use this number for our example.

Let's say that you are a consultant working for a fledgling company that is planning to make scrubs for nurses. Let's call this company Carol's Nursing Scrubs, Inc. Scrubs at Carol's come in small, medium, and large. The company offers all kinds of styles and prints, but the underlying sizes are intended to remain the same. Carol just received her first bit of seed money to mass-produce 20,000 pairs of scrubs. Carol, an overly demanding boss, wants the medium-size scrubs to fit as many nurses nationwide as possible. To make that happen, she needs to know the average height and weight of nurses nationwide, so she has instructed you to conduct a nationwide poll. She thinks you should ask every nurse in the country his or her height and weight and then calculate the average of all the numbers you get.

Now, you are an intelligent, well-grounded employee who's in demand everywhere and working for Carol only because her health plan comes with a sweet gym membership and you get a company car. You realize it would be pretty difficult to set up a nationwide poll and ask all the nurses in the country for their height and weight. Even if you tried a mass mailing, the data returned to you would be filled with so many incompletes and errors that it wouldn't be trustworthy.

So what are you to do? Your first instinct might be to respond to your boss by saying, "Geez, Carol, that's so absurd and impossible I don't even know where I'd start," and then finish your day on the golf range. After this course, however, you'll be not only a nurse but a nurse with some training in statistics. You'll be able to deal with this situation more effectively.

Jenna the Statistical Nursing Guru (you):	Carol, I recommend we take a few *samples* of nurses nationwide and *survey* them rather than attempting to contact every nurse in the country. Then we could *estimate* the true average height and weight based on our samples.
Carol:	How would that work, Jenna?
Jenna:	Well, I'd go down to the University Hospital and poll 30 RNs on their height and weight. Then I'd go to the next state and do the same. My third and final sample would contain 30 RNs from a hospital in Springfield. I'd calculate the average from my total *sample* (90 RNs), which is a *statistic*, and use that to estimate the overall average in the United States, which is a *parameter* of the total *population*. You see, Carol, anytime you calculate an estimate with data from a sample or list the data from the sample itself, you calculate a statistic. If you calculate an estimate from data in an entire population, you're calculating a parameter.

the patient's marital status (married, divorced, separated, living together, and other) is another example of nominal data.

Your next survey question asks the patient to rate his or her family support level as low, medium, or high. This question is an example of **ordinal data**. Ordinal data must be exhaustive and exclusive, just like nominal data, but the answers are also rank-ordered. With rank-ordered data, each observation/category is higher or lower, or better or worse, than another, but you do not know the level of difference between the observations/categories. In this example, a high level of family support indicates a greater quantity of the variable in question than does a moderate or low level of family support

A routine part of admitting each patient also includes a baseline set of vital signs, which you want to include in your survey data. One of the vital signs you check is each subject's temperature. Temperature is an example of **interval data**, which is exhaustive, exclusive, and rank-ordered, and has numerically equal intervals. In this example, the interval is a degree of Fahrenheit. You may also decide you want to look at your survey data by age group and develop the table shown in **Table 1-1**. In this example, age group is interval data; it is exhaustive, exclusive, and rank-ordered; and each interval is 4 years, so the intervals are all equal.

TABLE 1-1 **Number of Patients Surveyed in Each Age Group**

Age Group	Number of Patients
40–44 years	4
45–49 years	22
50–54 years	48
55–59 years	84

After recording each patient's temperature, you go on to examine each patient's blood pressure. Blood pressure is an example of **ratio data**, which is exhaustive, exclusive, and rank-ordered, with equal intervals *and* a point at which the variable is absent. (If the blood pressure reading is "absent" in any of your patients, you need to begin CPR!)

If you look at the diagram in **Figure 1-6**, you will see the relationship between the levels of measurement. Each increase in level includes the factors of the previous level, plus it adds another qualifier. Thus, if a variable is at the ratio level, it meets all the criteria for the nominal, ordinal, and interval levels, plus there is a point where it does not exist.

Ratio data is the highest level of measurement you can collect and gives you the greatest number of options for data analysis, but not all variables can be measured at this level. As a general rule of thumb, always collect the highest-level data you can for all your variables, especially your dependent variable. In your study of how family support (the independent variable) affects the duration of hospital stay (the dependent variable), you could have measured the length of hospital stay as short, medium, or long (ordinal) or in the number of actual days (the interval/ratio level). Obviously, the actual number of days gives you a higher level of measurement. A dependent variable with a higher level of measurement allows for a more robust data analysis. So collect the highest level you can! (See **Figure 1-7**.)

Note: Ordinal data may be quantitative (age group) or qualitative (mild/moderate/severe).

FIGURE 1-6 **Relationship between the Levels of Measurement.**

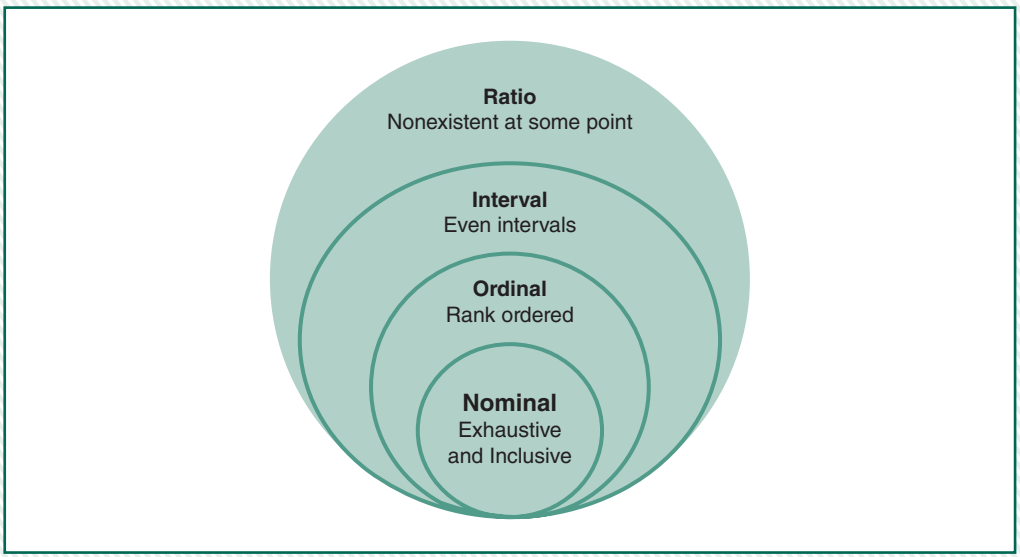
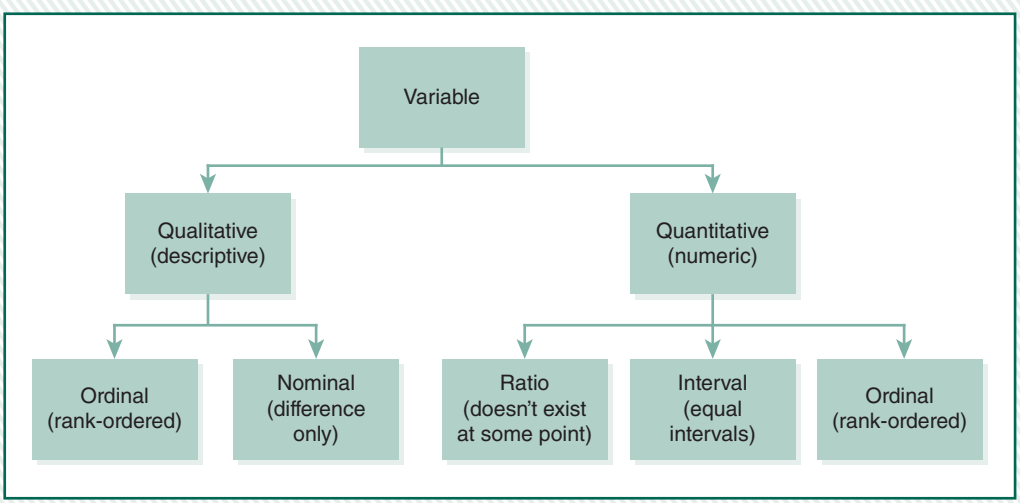


FIGURE 1-7 **The Relationship between Variable Descriptions.**



THINK IT THROUGH

How Can I Determine the Level of Measurement of a Variable?

Your study examines placental weight, which is measured in grams. What level of measurement is this variable?

Nominal Level: Ask yourself, does the variable show a difference? Yes, it does, different scores indicate different placental weight. This variable is at least at the nominal level.

If your answer is yes, then go to the next step because the variable may be at a higher level.

Ordinal Level: Is the difference rank-ordered? Yes, a lower score means less placental weight, and a higher score means more placental weight. Your variable is at least at an ordinal level.

If your answer to this question is no, then you do not have the criteria for this level and should identify your variable as the level before, in this case, nominal. If your answer is yes, then go to the next step.

Interval: Does this variable have equal intervals? Yes, each gram is an equal interval. Then your variable is at least interval level.

If your answer to this question is no, then you do not have the criteria for this level and should identify your variable as the level before, in this case, ordinal. If your answer is yes, then go to the next step.

Ratio: Is there a point where the variable can be equal to zero? No, every placenta will have at least some level of mass to it, so the variable will never be equal to zero.

Because your answer to this question is no, you do not have the criteria for this level and should identify your variable as the level before, in this case, interval. If your answer is yes, then you have satisfied the criteria for the highest level of measurement, which is ratio.

Let's look at another example, using the same steps. Your study examines amniotic fluid volume (AFV) measured as minimal, adequate, and excessive. What level measurement is this variable?

Nominal Level: Ask yourself, does the variable show a difference? Yes, it does; different categories indicate different amounts of AFV. This variable is at least at the nominal level.

If your answer is yes, then go to the next step because the variable may be at a higher level.

Ordinal Level: Is the difference rank-ordered? Yes, a person with minimal AFV has less AFV than someone with excessive AFV. Then your variable is at least at the ordinal level.

If your answer to this question is no, then you do not have the criteria for this level and should identify your variable as the level before, in this case, nominal. If your answer is yes, then go to the next step.

Interval: Does this variable have equal intervals? No, we don't know the intervals for these categories, so we can't say they are even.

Because your answer to this question is no, you do not have the criteria for this level and should identify your variable as the level before, in this case, ordinal.

Summary

Talk about exhausting, but you survived! So let's wrap it up here. Statistics really boils down to asking:

- Is what you observe what you expect?
- Or, using the empirical method, have you determined that what you observe is different enough from what you would expect that you need to change your expectations?

Using qualitative (descriptive) and quantitative (numeric) variables, you can assess the impact of independent variables on dependent (outcome) variables. Always collect the highest level of measurement possible, especially for your dependent variable. Doing so gives you the widest range of analysis options when you are ready to "crunch the numbers."

If you understand these concepts, you are ready to move on to the review exercises. If you are still struggling, don't despair. These concepts sometimes take a while to absorb. Read the review questions and then the chapter again, and slowly start to look at the review questions. You will get the hang of statistics; sometimes you just need practice. My students frequently look at me as though I am an alien when I tell them that by the end of the course this chapter will seem really simple. You may not believe it either. As you develop your understanding and apply these concepts, however, they will become clearer, and you too will look back in amazement. You are a statistical genius in the making!

CHAPTER 1 REVIEW QUESTIONS

1. A researcher asks hospitalized individuals about their comfort in a new type of hospital gown. This is an example of what type of data?	a. ratio b. independent c. quantitative d. qualitative
2. If a researcher is examining how exposure to cigarette ads affects smoking behavior, the cigarette ads are what type of variable?	a. qualitative b. quantitative c. dependent d. independent
3. A nurse practitioner measures how many times per minute a heart beats when an individual is at rest versus when running. She is measuring the heartbeat at what level of measurement?	a. interval/ratio b. nominal c. independent d. ordinal
4. If a researcher is examining how exposure to cigarette ads affects smoking behavior, smoking behavior is what type of variable?	a. ratio b. independent c. dependent d. nominal
5. The research nurse is coding adults according to size. A person with a below-average body mass index (BMI) is coded as 1, average is 2, and above average is 3. What level of measurement is this?	a. nominal b. ratio c. ordinal d. interval

6. You are asked to design a study measuring how nutritional status is related to serum lead levels in children. You assess calcium and fat intake, as well as serum lead levels in a sample of 30 children who are 2 years old. Lead levels are measured in micrograms per deciliter (mcg/dL). One child had a lead level of 17 mcg/dL. This is an example of what type of variable?
   a. quantitative
   b. qualitative
   c. independent
   d. nominal

Questions 7–9: You are asked to design a study to examine the relationship between preoperative blood pressure and postoperative hematocrit.

7. What is your independent variable?

8. What is your dependent variable?

9. How will you measure each, and what level of measurement is this?

Questions 10–13: You are later asked to do a follow-up study to see whether requiring an intraoperative blood transfusion had an impact on postoperative rates of poor mental health, specifically depression.

10. What is your independent variable?

11. What is your dependent variable?

12. How will you measure these variables, and why?

13. Is your dependent variable measured at the highest level? If not, why not?

Questions 14–18: You decide to measure depression on the following scale: 1 = low, 2 = moderate, 3 = high.

14. What level of measurement is this?

15. How could this measure be improved?

16. Why might you want to improve it?

17. You decide to measure postoperative hematocrit by serum levels. Is this a quantitative or qualitative measurement?

18. You discover that all but those with the lowest hematocrits had higher levels of depression after their surgery and transfusion. Why might the group that had the most critical need for the transfusions not have the subsequent depression associated with this result in the rest of your sample?

Questions 19–25: Elevated serum lead levels in childhood are associated with lower IQ, hyperactivity, aggression, poor growth, diminished academic performance, increased delinquency, seizures, and even death. The neurological damage that occurs cannot be reversed, even once exposure is stopped.

19. You have been asked to follow up in your community and determine what outcomes are associated with lead exposure in children. List three dependent variables for your study and how you will measure them.

20. What level of measurement are your dependent variables? Are they continuous or categorical?

21. Can you increase the level of measurement for any of them?

22. If you are looking at what outcomes are associated with lead exposure in children, what is your independent variable?

23. Why might this independent variable be difficult to measure?

24. Describe how this independent variable could be measured quantitatively or qualitatively.

25. Which way do you prefer to measure the independent variable? Why?

Questions 26–34: A nurse researcher is assessing how well patients respond to two different dosing regimens of a new drug approved to treat diabetic neuropathy. Two different dosing regimens are administered, and side effects are monitored. Results are shown in **Table 1-2**.

TABLE 1-2 **Self-Reported Side Effects of Two Randomized Groups of 100 Individuals Treated for Diabetic Neuropathy**

Side Effect Reported	Low Dosage	High Dosage
Nausea	8	21
Headache	3	5
Weight gain	1	0
Weight loss	0	6
Lethargy	3	11
Skin rash	13	13

26. What is the independent variable?

27. What is the dependent variable?

28. In this study, the nurse researcher measures the side effects as present or not present. This variable is what level of measurement?

29. If instead the nurse researcher decided to measure weight gain in pounds, what level of measurement would it be? Would it be a continuous or categorical variable?

30. If the nurse researcher decided to measure nausea as present, limiting, or debilitating, what level of measurement would it be? Would nausea be a continuous or categorical variable?

31. If the nurse researcher measured nausea as the number of hours of nausea experienced in a day, what level of measurement would it be?

32. If the nurse researcher asked the subjects to describe their headache, would this be a quantitative or a qualitative variable?
In the second phase of this study, the nurse researcher asks the study participants to report changes in signs and symptoms of their neuropathy. She determines that those on the low-dose regimen had a similar level of pain relief and improvement in mobility as those who took the high-dose drug regimen.

33. What is the dependent variable in the second phase of the study?

34. Considering the information you now know about the side effects and relief of neuropathy symptoms, what might you prefer as a patient? Why? What else might you want to know before making the decision?

Questions 35–38: Relate to the following content.

35. You complete a study in which you categorize the subject's blood pressure as normal, prehypertensive, high blood pressure stage 1, or high blood pressure stage 2 using the following criteria. What level of measurement is the stage of high blood pressure?
36. You are now interested in examining compliance with a DASH diet. You ask your subjects if they have or have not complied with the diet this week. Your dietary compliance variable is what level of measurement?
37. After meeting with your statistician, you measure compliance with the DASH diet on a scale of 1 to 7. For analysis purposes, the dietary compliance variable is now what level of measurement? Why might the statistician have recommended this change?
38. You conclude your study by examining how compliance with the DASH diet affects the stage of high blood pressure. What is your independent variable?
39. What is your dependent variable? Is it continuous or categorical?

Long-range winter forecast for 2017

Modified from ©Cartoon Resource/Shutterstock.

ANSWERS TO ODD-NUMBERED CHAPTER 1 REVIEW QUESTIONS

1. d
3. a
5. c
7. Preoperative blood pressure
9. Answers will vary: actual blood pressure ratio, lab-reported hematocrit ratio, and so on.
11. Depression
13. Answers will vary.
15. Use of interval data, such as Beck's depression scale
17. Quantitative
19. Answers will vary: including IQ, school enrollment, crime, pregnancy, hematocrit, learning disabilities, growth, hearing, and behavior.
21. Answers will vary.
23. Answers will vary: including "It requires a blood draw," "There are different testing mechanisms," "The level may change depending on when the exposure occurred and the time that has lapsed since then," "Levels may differ from fingersticks versus serum draws."
25. Answers will vary.
27. Side effects, nausea, headache, weight gain, weight loss, lethargy, skin rash
29. Ratio, continuous
31. Ratio
33. Signs and symptoms of neuropathy
35. Ordinal
37. Interval; because collecting the data at a higher level of measurement gives you more analysis options
39. Stage of high blood pressure, categorical