

Unit 1

Variables

Objectives:

- To decide whether a variable should be treated as quantitative, qualitative-ordinal, qualitative-nominal, or qualitative-dichotomous
- To understand the importance of validity, unbiasedness, and reliability when using an instrument to make observations of a variable

How would you define *statistics*? Whatever your definition of statistics might be, it would surely involve numbers. We generally use numbers to measure something. However, there is much more to the field of statistics than simply obtaining numbers to measure something. Statistics involves collecting information, organizing the information, and interpreting the information.

Whenever we collect information about objects or people, we are observing some characteristic of each object or person. For instance, we might want to obtain information about a person's eye color, or about the amount of chocolate in a chocolate chip cookie, or about how well a medication relieves cold symptoms. Any characteristic which is potentially different when measured for different items or at different times is called a variable.

There are two types of variables we can measure. A *quantitative variable* is one which we measure with meaningful numerical values; a *qualitative variable* is one which we measure with categories possessing no meaningful numerical values. Sometimes it is very clear whether a variable should be treated as quantitative or qualitative, but sometimes this is not immediately obvious.

For instance, suppose the variable "eye color" is of interest. A person's eye color would most likely be placed into one of several categories, such as blue, green, brown, hazel, etc. (with the number of categories determined by how much refinement in distinction between colors is desired). Clearly, eye color must be treated as qualitative, since no meaningful numerical values can be involved in any way. Note that for the sake of convenience we may want to use numerical values to code the categories; for example, a one (1) could be used to represent the color blue, a two (2) could be used to represent the color green, a three (3) could be used to represent the color brown, etc., but the value of the numbers have no meaningful connection to the different color categories.

Now suppose the variable "amount of chocolate in a chocolate chip cookie" is of interest. If we choose place each cookie into categories such as a little chocolate, a fair amount of chocolate, or a lot of chocolate, then we would be treating this variable as qualitative, since no meaningful numerical values are involved in any way. On the other hand, if we measured this variable by the weight of chocolate in grams for each cookie, then we would be treating this variable as quantitative, since meaningful numerical values are used. If we measured this variable by counting the number of chocolate chips in each cookie, we would also be using meaningful values and thus treating the variable as quantitative.

Some variables have no natural way of being measured. While it is quite natural to measure weight using grams, age using years, income using dollars, and number of children by a head count, there is no natural way to measure variables such as job satisfaction, intelligence, pain, and happiness. When no natural way to measure a variable exists, a method to measure the variable has to be developed.

Suppose the variable "how well a medication relieves cold symptoms" is of interest. There is no natural numerical method of measurement available. One possibility is to classify people's reaction to the medication into five categories labeled very poor, poor, slightly effective, moderately effective, and very effective. We would then be treating the variable as qualitative, since we have no meaningful numerical value associated with each category. On the other hand, we might attempt to measure "how well a medication relieves cold symptoms" as a quantitative variable by assigning values to each of the five categories; we might let 1 represent very poor, 2 represent poor, 3 represent slightly effective, 4 represent moderately effective, and 5 represent very effective. Alternatively, we might use a rating from 0 to 10, where 0 represents no effect and 10 represents extremely effective. If we treat the assigned values as if they are meaningful numerical values, then we are treating the variable as quantitative.

Self-Test Problem 1-1. Consider each variable in the SURVEY DATA, displayed as Data Set 1-1 at the end of this unit.

- (a) Decide which variables should clearly be treated as quantitative.
- (b) Decide which variables should clearly be treated as qualitative.
- (c) Decide which variables could potentially be treated as either quantitative or qualitative.

Self-Test Problem 1-2. Decide which of the following are variables and which are not:

- (a) a person's sex,
- (b) a female,
- (c) a 12-ounce box of cereal,
- (d) the amount of cereal in a box.

We can identify three types of qualitative variables. Qualitative variables having only two categories are called *qualitative-dichotomous*. The variable "whether or not a light bulb works" has only two categories, which we can label good and burned out. For qualitative variables having more than two categories, it may or may not be possible to put the categories in a natural order. When the categories have no natural ordering, the qualitative variable is called *qualitative-nominal*. Recall that the variable "eye color" measured with categories blue, green, brown, etc. is qualitative. Since the colors have no natural ordering, "eye color" is treated as qualitative-nominal. When the categories have a natural ordering, the qualitative variable is called *qualitative-ordinal*. Recall that the variable "how well a medication relieves cold symptoms" measured with categories very poor, poor, slightly effective, moderately effective, and very effective is qualitative. Since these categories have an obvious natural ordering, "how well a medication relieves cold symptoms" could be treated as qualitative-ordinal.

Self-Test Problem 1-3. For each of the following ways of measuring the variable "Income," decide whether the variable can be treated as quantitative, qualitative-ordinal, qualitative-nominal, or qualitative-dichotomous:

- (a) the amount of dollars of income per year;
- (b) the income class to which a person belongs labeled as "Low Income," "Middle Class," or "High Income";
- (c) the source(s) of income for a person.

Self-Test Problem 1-4. For each of the following ways of measuring the variable "Bus Riding Habits," decide whether the variable can be treated as quantitative, qualitative-ordinal, qualitative-nominal, or qualitative-dichotomous:

- (a) identifying which of the 20 busses that run each weekday, a person rides;
- (b) classifying how often a person rides the busses as "never," "occasionally," "one or two days a week," or "every day";
- (c) recording the number of times a person rides the busses in a given week;
- (d) recording the number of times a person rides the busses in a given month.

Self-Test Problem 1-5. In parts (b) and (c) of Self Test Problem 1-1, you found that some of the variables in the SURVEY DATA could be treated as qualitative. Decide whether each of these variables is qualitative-dichotomous, qualitative-nominal, or qualitative-ordinal.

Earlier we observed that some variables, such as job satisfaction, intelligence, pain, and happiness, have no natural way of being measured, while other variables have a natural way of being measured, such weight in grams, age in years, income in dollars, and number of children by a head count. When no natural way to measure a variable exists, a method to measure the variable has to be developed.

In statistical terminology, an *instrument* refers to the mechanism we use to measure a variable. For instance, to obtain a person's weight directly, we could ask each person to step onto a scale. On the other hand, we may choose not to measure weight directly, but instead we may simply ask each person for his or her weight as part of either a verbal or written questionnaire. In the first case, the scale would be our instrument, but in the second case, the questionnaire would be our instrument.

Validity refers to the appropriateness of an instrument used to measure a variable. The validity of measuring weight by using a scale would certainly be perfectly appropriate, since measuring weights is exactly what a scale is designed to do. However, the validity of measuring weight by simply ask each person for his or her weight could be challenged on the basis that not all people may have weighed themselves recently and consequently may not really be able to estimate their weight reasonably. To decide whether asking a person for his or her weight is valid measure of weight, we could simply compare responses with actual weight measurements using a scale (assuming of course that respondents agreed to be weighed).

One could question the validity of measuring “how well a medication relieves cold symptoms” by simply asking each person on the basis that expressions such as very effective or moderately effective might not mean exactly the same thing to all people. It is not as easy to verify validity as it might be to verify the validity of measuring weight by simply ask each person for his or her weight; this is because we could compare stated weights with actual weights, but there is no natural measurement with which to compare a person’s responses about how well a medication relieves cold symptoms. As a general rule, we must always be concerned about validity whenever we must devise a way to measure a variable for which there is no natural method of measurement. This is typically the case when a questionnaire or test is designed to measure variables such as satisfaction, intelligence, pain, and happiness. Often, the only way to verify the validity of a questionnaire is to compare responses to those on other questionnaires which are known to be valid and measure the same or closely related variables.

Once we are convinced that an instrument is valid, we can turn our attention to the accuracy of measurements. For instance, when we use a scale to measure the variable weight, we do not question the validity, but we may certainly be concerned with the accuracy of our measurements. Two aspects of accuracy need to be considered. The measurement of a variable is called *biased* if there is a tendency for errors to occur in one direction; the measurement is called *unbiased* if the value of observations is correct on average with independent, repeated measurements on the same item. The measurement of a variable is called *unreliable* if there is a tendency for errors to occur in any direction; the measurement is called *reliable* if the value of observations is the same, or approximately the same, with independent, repeated measurements on the same item.

A scale which consistently displays eight pounds higher than the correct weight, or radar equipment used to clock the speed of cars which consistently measures ten miles per hour below the correct speed, or a survey question which is slanted to elicit a response in a particular direction, are all examples of bias. In order to be unbiased, the scale would have to display correct weights on average, the radar equipment would have to measure the correct speed of cars on average, and the survey question would have to be worded so that the reader does not get the impression that a certain response is desired.

Even if measurements are unbiased, the measurements would be considered unreliable if there were drastically different results with independent, repeated measurements of the same item. For instance, suppose a scale is used to obtain five independent weight readings of an object weighing exactly 50 grams. If the five weight readings were found to be 45.09, 54.92, 49.97, 46.95, and 53.05, then we then we would say that the scale is unreliable because of the widely different measurements of the same object, even though it would also seem that the scale is unbiased, since the average of the five readings is the correct weight of 50. On the other hand, if the five readings were found to be 51.09, 50.92, 50.97, 50.95, and 51.05, then we could say that the scale is reliable because of the relatively close agreement among measurements of the same object, but it would also seem that the scale is biased, since the average of the five readings is almost 1 gram higher than the correct weight of 50. Finally, if the five weight readings were found to be 50.09, 49.92, 49.97, 49.95, and 50.05, we could still say that the scale is reliable because of the relatively close agreement among measurements of the same object, but it would also seem that the scale is unbiased, since the average of the five readings is relatively close to the correct weight of 50.

To summarize, then, we can say that bias refers to errors in measurement that are consistently made in one direction, while unreliability refers to errors which occur in both directions. It may be possible to correct measurements for bias, if we know something about what the direction and size of the bias are likely to be. However, it is never possible to correct for unreliability. Finally, we must always be certain that the way we measure a variable is valid before considering unbiasedness or reliability.

Self-Test Problem 1-6. Decide whether each of the following addresses the issue of validity, the issue of unbiasedness/bias, or the issue of reliability/unreliability:

- (a) A gauge is used to measure the air pressure inside a bike tire several times in succession to see how closely the readings agree with each other.
- (b) A questionnaire designed to be completed by a student is used to measure an instructor's performance. The numerical score generated by the questionnaire is criticized as measuring the instructor's popularity instead of the instructor's performance.
- (c) Decide Two different gauges are each used to measure the air pressure inside a bike tire. One gauge has a brand name with an excellent reputation for accuracy while the other gauge has brand name which is unfamiliar. It is found that measurements from the gauge with the unfamiliar brand name seem to consistently be about 5 or 6 pounds per square inch less than those with the other gauge.

Self-Test Problem 1-7. There is no natural way to measure the variable pain. For each of the following ways of measuring pain, decide whether pain would most likely be treated as quantitative, qualitative-ordinal, qualitative-nominal or qualitative-dichotomous:

- (a) Subjects are asked to describe pain as either mild or intense.
- (b) Subjects are asked to choose from a list of several adjectives to describe the type of pain.
- (c) Subjects are asked to choose from a list of several adjectives to describe the intensity of pain.
- (d) Subjects are asked to choose from a list of adjectives to describe the type of pain, where each adjective is labeled with an integer from 1 to 7.
- (e) Subjects are asked to rate the level of pain by selecting a point on a line segment where one end of the line segment is labeled 0 to represent no pain and the other end of the line segment is labeled 1 to represent unbearable pain.

Self-Test Problem 1-8. A study is being conducted to determine the knowledge that sixth-graders possess about nutrition.

- (a) What type of instrument might be used to measure the knowledge of a sixth-grader about nutrition, and how might the validity of the instrument be verified?
- (b) Suppose an exam consisting of 30 questions is constructed to measure the knowledge of a sixth-grader about nutrition; the number of questions answered correctly is the score used to measure knowledge. Suppose this score is a valid measure of knowledge and is administered twice to a sixth grader who is knowledgeable about only half of the material covered by the exam. If the two scores were observed to be 14 and 16, what might this suggest about the unbiasedness and reliability of the exam?
- (c) Suppose the score described in part (b) is a valid measure of knowledge and is administered twice to a sixth grader who is knowledgeable about only half of the material covered by the exam. If the two scores were observed to be 10 and 11, what might this suggest about the unbiasedness and reliability of the exam?
- (d) Suppose the score described in part (b) is a valid measure of knowledge and is administered twice to a sixth grader who is knowledgeable about only half of the material covered by the exam. If the two scores were observed to be 10 and 20, what might this suggest about the unbiasedness and reliability of the exam?
- (e) Suppose the score described in part (b) is a valid measure of knowledge and is administered twice to a sixth grader who is knowledgeable about only half of the material covered by the exam. If the two scores were observed to be 5 and 13, what might this suggest about the unbiasedness and reliability of the exam?

Answers to Self-Test Problems

- 1-1** (a) "Number of Children," "Age," "Yearly Income," "Weekly TV Hours," and "Weekly Radio Hours" are each obviously treated as a quantitative variable, since each is measured with meaningful numerical values. (b) "ID No.," "Sex," "Residence," and "Political Party Affiliation" are each treated as a qualitative variable, since each of these variables is measured with categories possessing no meaningful numerical values. (c) "Job Satisfaction Score" could be treated as qualitative if the numerical values were considered to be codes with no meaning, or could be treated as quantitative if the numerical values were interpreted to have meaning.
- 1-2** (a) a variable (b) not a variable (c) not a variable (d) a variable
- 1-3** (a) quantitative (b) qualitative-ordinal (c) qualitative-nominal
- 1-4** (a) qualitative-nominal (b) qualitative-ordinal (c) quantitative (d) quantitative
- 1-5** "Sex" is treated as qualitative-dichotomous. "ID No." and "Political Party Affiliation" are each qualitative-nominal. "Residence" could be treated as qualitative-ordinal, since the three categories could be ordered with rural at one end and urban at the other end.
- 1-6** (a) reliability/unreliability (b) validity (c) unbiasedness/bias
- 1-7** (a) qualitative-dichotomous (b) qualitative-nominal (c) qualitative-ordinal (d) quantitative (e) quantitative
- 1-8** (a) One possibility is to administer an exam consisting of questions about nutrition. The validity of such an exam might be verified by administering the exam together with another instrument known to be a valid measure of nutrition knowledge, or a similar variable, to see how strongly the scores from the two instruments are related.
(b) The score is unbiased and reliable. (c) The score is reliable but biased.
(d) The score is unbiased but unreliable. (e) The score is biased and unreliable.

Summary

A characteristic which is potentially different when measured for different items or at different times is called a *variable*. A variable is called *quantitative* when it is measured with meaningful numerical values. A variable is called *qualitative-ordinal* when it is measured with categories that can be ordered in a meaningful way. A variable is called *qualitative-nominal* when it is measured with categories that cannot be ordered in a meaningful way. Variables having only two categories are called *qualitative-dichotomous*.

Validity refers to the appropriateness of an instrument used to measure a variable. We must always be certain that the way we measure a variable is valid before considering unbiasedness or reliability. The measurement of a variable is called *biased* if there is a tendency for errors to occur in one direction; the measurement is called *unbiased* if the value of observations is correct on average with independent, repeated measurements on the same item. The measurement of a variable is called *unreliable* if there is a tendency for errors to occur in any direction; the measurement is called *reliable* if the value of observations are the same, or approximately the same, with independent, repeated measurements on the same item. It may be possible to correct measurements for bias, if we know something about what the direction and size of the bias are likely to be. However, it is never possible to correct for unreliability.

Data Set 1-1 SURVEY DATA

Voters in a particular state are surveyed. Each respondent is assigned an identification number, and information about each of the following is recorded: sex of the voter, area of residence, political party affiliation, number of children, yearly income, job satisfaction, weekly hours spent watching TV, and weekly hours spent listening to radio. Job satisfaction is measured by having each voter give a rating from 0 to 10, where 0 represents totally dissatisfied and 10 represents totally satisfied.

ID No.	Sex	Residence	Political Party Affiliation	Number of Children	Age	Yearly Income (\$1000s)	Job Satisfaction Score	Weekly TV Hours	Weekly Radio Hours
08	M	Suburban	Republican	5	35	34	5	15	15
27	F	Urban	Democrat	2	20	28	2	20	13
34	F	Suburban	Independent	4	35	71	7	18	17
18	M	Rural	Independent	7	41	35	8	12	20
04	M	Urban	Republican	3	39	55	4	14	15
14	M	Urban	Democrat	3	59	75	1	11	18
23	F	Urban	Democrat	1	20	26	2	10	11
39	M	Rural	Other	4	52	30	9	12	30
54	F	Rural	Other	2	44	27	7	8	14
44	M	Urban	Republican	0	46	53	0	21	10
17	F	Urban	Republican	2	40	45	3	12	14
12	F	Suburban	Other	2	34	34	5	8	10
26	F	Urban	Republican	2	24	30	2	11	15
11	M	Urban	Democrat	4	62	78	1	18	11
38	M	Suburban	Other	3	44	68	7	17	12
09	F	Rural	Republican	6	44	29	9	8	27
29	F	Suburban	Democrat	4	38	40	9	9	25
13	M	Rural	Independent	9	47	39	6	8	20
24	M	Urban	Democrat	3	44	60	2	27	4
33	F	Urban	Democrat	1	45	49	3	9	12
15	F	Rural	Other	2	56	39	8	6	27
52	F	Suburban	Republican	0	32	33	5	11	15
35	M	Suburban	Other	1	54	65	3	14	14
30	F	Rural	Independent	3	41	25	7	5	25
47	M	Suburban	Republican	6	50	61	5	15	15
32	F	Rural	Republican	2	59	41	8	8	27
41	F	Suburban	Other	3	44	44	3	10	18
10	M	Rural	Other	3	62	45	10	10	23
48	M	Suburban	Republican	2	53	64	1	13	15
02	M	Rural	Democrat	8	59	39	8	9	24