Solution *Section* 1.1 – Statistical Thinking

Exercise

Use common sense to determine whether the given event is (***a***) *impossible*; (***b***) *possible, but very unlikely*; (***c***) *possible and likely*.

a) Giants best the Denver Broncos in the Super Bowl by a score of 120 to 98.

b) While driving to his home in Connecticut, David was ticketed for driving 205 *mi/h* on a highway with a speed limit of 55 *mi/h*.

c) Thanksgiving Day will fall on a Monday nest year.

d) When each of 25 statistics students turns on his or her TI-84 Plus calculator, all 25 calculators operate successfully.

Solution

a) (***b***) Possible, but very unlikely.
Possible because there are no physical constraints or rules of the game that would prohibit such scores.

b) (***b***) Possible, but very unlikely.
Possible if he were driving a race car that could attain such a speed.

c) (***a***) Impossible, In the U.S., Thanksgiving Day always falls on a Thursday.

d) (***c***) Possible and likely

Exercise

Determine whether the underline value is a ***parameter*** or a ***statistic***.

a) Following the 2010 national midterm election, 18% of the governors of the 50 United States were female.

b) The average score for a class of 28 students taking a calculus midterm exam was 72%.

c) In a national survey of 1300 high school students (grades 9 to 12), 32% of respondents reported that someone has bullied them at school.

d) In a national survey on substance abuse, 10.0% of respondents aged 12 to 17 reported using illicit drugs within the past month.

e) Ty Cobb is one of major league baseball's greatest hitters of all time, with a career batting average of 0.366.

f) Only 12 men have walked on the moon. The average age of these men at the time of their moonwalks was 39 years, 11 months, 15 days.

g) A study of 6076 adults in public rest rooms (in Atlanta, Chicago, New York City, and San Francisco) found that 23% did not wash their hands before exiting.

h) Interviews of 100 adults 18 years of age or older, conducted nationwide, found that 44% could state the minimum age required for the office of U.S. president.

i) In a large sample of households, the median annual income per household for high school graduates is $19,856 (based on data from the U.S. Census Bureau).

j) Among the Senators in the current Congress, 44% are Democrats.

k) A study of all 2223 passengers aboard the Titanic found that 706 survived when it sank.
l) If the areas of the 50 states are added and the sum is divided by 50, the result is 196,533 square kilometers.
m) The average (mean) atomic weight of all elements in the periodic table is 134,355 unified mass units

Solution

a) 18% is a ***parameter*** because it describes a population (all of the governors).

b) 72% is a ***parameter*** because it describes a population (the entire class).

c) 32% is a ***statistic*** because it describes a sample (the high school students surveyed).

d) 10% is a ***statistic*** because it describes a sample (the youths surveyed).

e) 0.366 is a ***parameter*** because it describes a population (all Ty Cobb's at-bats).

f) 39 yrs, 11 mths, 15 days is a ***parameter*** because it describes a population (all Ty Cobb's at-bats).

g) 23% is a ***statistic*** because it describes a sample (the 6,076 adults studied).

h) 44% is a ***statistic*** because it describes a sample (the 100 adults interviewed).

i) ***Statistic***, since it was determined from a sample of households.

j) ***Parameter***, since it is based on the population of all Senators.

k) ***Parameter***, since it was determined from the population of all 2223 passengers.

l) ***Parameter***, since it was determined from the population of all 50 states.

m) ***Parameter***, since it was determined from the population of all elements in the periodic table.

Exercise

Classify the variable as ***qualitative*** or ***quantitative***

a) Nation of origin
b) Number of siblings
c) Grams of carbohydrates in a doughnut
d) Number on a football player's jersey
e) Number of unpopped kernels in a bag of ACT microwave popcorn
f) Assessed value of a house
g) Phone number
h) Student ID number.
i) Favorite film
j) Population of country of origin
k) Gallons of water in a swimming pool
l) Model of car driven
m) Distance in miles to nearest school
n) Time in hours that a light bulb lasts
o) Number of students at a high school

Solution

a) ***Qualitative***

b) Quantitative

c) Quantitative

d) Qualitative

e) Quantitative

f) Quantitative

g) Qualitative

h) Qualitative

i) Qualitative

j) Quantitative

k) Quantitative

l) Qualitative

m) Quantitative

n) Quantitative

o) Quantitative

Exercise

Determine whether the quantitative variable is ***discrete*** or ***continuous***

a) Goals scored in a season by a soccer player
b) Volume of water lost each day through a leaky faucet
c) Length (in minutes) os a country song
d) Number of Sequoia trees in a randomly selected acre of Yosemite National Park
e) Temperature on a randomly selected day in Memphis, Tennessee
f) Internet connection speed in Kilobytes per second
g) Points scored in an NCAA basketball game
h) Air pressure in pounds per square inch in an automobile tire
i) In the Literary Digest poll, Landon received 16,679,583 votes
j) The amount of nicotine in a Marlboro cigarette is 1.2 mg
k) The volume of cola in a can of regular coke is 12.3 oz
l) When a woman is randomly selected and measured for blood pressure, the systolic blood pressure is found to be 61 mm Hg

Solution

a) Discrete

b) Continuous

c) Continuous

d) Discrete

e) Continuous

f) Continuous

g) Discrete

h) Continuous

i) Discrete, since the number of votes received must be a whole number

j) Continuous, since the amount of nicotine could be any value on a continuum; even though it is reported to the nearest 0.1 mg.

k) Continuous, since the volume could be any value on a continuum; even though it is reported to the nearest 0.1 oz.

l) Continuous, since the systolic pressure could be any value on a continuum; even though it is reported to the nearest 61 mm Hg.

Exercise

Determine the level of measurement of each variable

a) Nation of origin
b) Movie ratings of one star through five stars
c) Volume of water used by a household in a day
d) Year of birth of college students
e) Highest degree conferred (high school, bachelor's, and so on)
f) Eye color
g) Assesses value of a house
h) Time of day measured in military time
i) Types of movies (drama, comedy, adventure, documentary, etc.)
j) Critic ratings of movies on a scale from 0 star to 4 stars
k) Ranks of cars evaluated by Consumer's Union

Solution

a) Nominal

b) Ordinal

c) Ratio

d) Interval

e) Ordinal

f) Nominal

g) Ratio

h) Interval

i) Nominal, since the data give category names only and there is no natural ordering.

j) Ordinal, since the categories have a natural ordering but the differences between the categories are not necessarily uniform.

k) Ordinal, since the ranks have a natural ordering but the differences between the ranks are not necessarily uniform.

Exercise

The Gallup Organization contacts 1026 teenagers who are 13 to 17 years of age and live in the United States and asks whether or not they had been prescribed medications for any mental disorders, such as depression or anxiety. Identify the population and sample.

Solution

The population consists of all teenagers 13 to 17 years old who live in the U.S.

The sample consists of the 1,028 teenagers 13 to 17 years old who were contacted by the Gallup Organization.

Exercise

A quality-control manager randomly selects 50 bottles of Coca-Cola that were filled on October 15 to assess the calibration of the filling machine. Identify the population and sample.

Solution

The population consists of all bottles of Coca-Cola filled by that particular machine on October 15.

The sample consists of the 50 bottles of Coca-Cola that were selected by the quality control manager.

Exercise

Each *x* value associated with the corresponding *y* value in some meaningful way? If the *x* and *y* values are not matched, does it make sense to use the difference between each *x* value and the *y* value that is the same column?

Nicotine Amounts from Menthol and King-Size Cigarettes

x	1.1	0.8	1.0	0.9	0.8
y	1.1	1.7	1.7	1.1	1.1

The *x*-values are nicotine amounts (in *mg*) in different 100 *mm* filtered, non-light menthol cigarettes; the *y*-values are nicotine amounts (in mg) in different king-size non-filtered, non-menthol, and non-light cigarettes.

Solution

There is nothing in the data to suggest that the *x* and *y* values are matched. If the *x* and *y* values are not matched as presented, it does not make sense to calculate the differences between the *x* and *y* values – moving *x* values around, for example, would yield a different set of differences.

Exercise

The Federal Trade Commission obtained the measured amounts of nicotine in the table. Is the source of the data likely to be unbiased?
Note that the table lists measured nicotine amounts from two different types of cigarette. Given these data, what issue can be addressed by conducting a statistical analysis of the values?

Nicotine Amounts from Menthol and King-Size Cigarettes

x	1.1	0.8	1.0	0.9	0.8
y	1.1	1.7	1.7	1.1	1.1

Solution

Yes, since the Federal Trade Commission receives no financial or other benefit from a particular set of results, there is no reason for the data or the data selected to be biased.

Exercise

One of Gregor Mendel's famous hybridization experiments with peas yielding 580 off spring with 152 of those peas (or 26%) having yellow pods. According to Mendel's theory, 25% of the off spring peas should have yellow pods. Do the results of the experiment differ from Mendel's claimed rate of 25% by an amount that is statistically significant?

Solution

Determining whether or not the difference between Mendel's actual results (26%) and the results predicted by his theory (25%) is statistically significant properly requires applying some techniques. Common sense suggests the 1% difference is of no practical difference. Considering the sample size, the actual difference between the observed and expected results is 152 – 145 = 7. Common sense suggests that a discrepancy of 7 (relative to an expected result of 145 plants from a total sample of 580 plants) is within the natural fluctuation inherent biological processes, and that the difference is not statistically significant.

Exercise

In a Gallup poll of 1038 randomly selected adults, 85% said that secondhand smoke is somewhat harmful or very harmful, but a representative of the tobacco industry claims that only 50% of adults believe that secondhand smoke is somewhat harmful or very harmful. Is there statistically significant evidence against the representative's claim? Why or why not?

Solution

Determining whether or not the difference between the survey results (85%) and the statement of the industry representative (50%) is statistically significant properly. But common sense suggests that this is statistically significant evidence against the representative's claim because (1) the observed result of 85% is so much greater than the stated claim of the 50% and (2) the sample of 1038 randomly selected adults appears to be properly chosen and large enough to provide reliable data.

Exercise

Determine whether the given value is parametric or a statistic

a) One of greatest baseball hitters of all time has a career batting average of 0.366
b) A sample of employees is selected and it is found that 50% own a vehicle
c) A survey of 42 out of hundreds in a dining hall showed that 17 enjoyed their meal

Solution

a) Parameter
b) Statistic because the value is a numerical measurement describing a characteristic of a sample
c) Statistic, because the data set of 42 people in a dining hall is a sample.

Exercise

Suppose a survey of 568 women in the U.S. found that more than 61% are the primary investor in their household.

a) Describe the survey represents the descriptive branch of statistic
b) Make an inference based on the results of the survey

Solution

a) 61% of women in the sample are the primary investor in their household
b) There is an association between U.S. women and being the primary investor in their household

Exercise

In the recent study, volunteers who had 8 hours of sleep were three times more likely to answer questions correctly on a math test than were sleep-deprived participaants.

a) Identify the sample used in the study
b) What is the sample's population
c) Which part of the study represents the descriptive branch of statistics
d) Make an inference based on the results of the study

Solution

a) The sample is the responses of the volunteers in the study
b) The population is the collection of the responses of all individuals who completed the math test
c) The statement "three times more likely to answer questions correctly" is an example of descriptive statistics
d) Individuals who are not sleep deprived will be more likely to answer math questions correctly than individuals who are sleep deprived.

Exercise

Determine whether the data set is a population or a sample. Explain your reasoning

The salary if each baseball player in a league

Solution

Population, because it is a collection of salaries for all baseball players in the league

Exercise

In a poll, 1,004 adults in a country were asked whether they favor or oppose the use of "federal tax dollars to find medical research using stem cells obtained from human embryos." Among the responders, 48% said that they were in favor. Describe the statistical study

a) What is the population?
b) Identify the sample

Solution

a) All adults in the country
b) The 1,004 adults selected

Exercise

A study shows that the obesity rate among boys ages 2 to 19 has increased over the past several years

a) Make an inference based on the results of this study?
b) What is wrong with this type of reasoning

Solution

a) The obesity rate among boys ages 2 to 19 is increasing
b) This inference may incorrectly imply that the trend will continue in future years.

Exercise

The newspaper USA Today published a health survey, and some readers completed the survey and returned it. Identify the (*a*) sample and (*b*) population, also determine whether the sample likely to be representative of the population.

Solution

a) The sample is the readers who completed and returned the survey.
b) The population is all people who read USA Today.
c) No, since the sample is self-selected it is not likely to be representative of the population. Only those with special interest in health matters are likely to return the survey.

Exercise

A Gallup poll of 1012 randomly surveyed adults found that 9% of them said cloning of humans should be allowed. Identify the (*a*) sample and (*b*) population, also determine whether the sample likely to be representative of the population.

Solution

a) The sample is the 1012 randomly selected adults.
b) The population is all adults. This should probably be understood to be all adults living in the U.S.
c) Yes, since the adults were selected at random by an organization with experience in polling and with no vested interest in the results, the sample is likely to be representative of the population.

Exercise

Some people responded to this request: "Dial 1-900-PRO-LIFE to participate in a telephone poll on abortion. ($1.95 per minute. Average call: 2 minutes. You must be 18 years old.)" Identify the (*a*) sample and (*b*) population, also determine whether the sample likely to be representative of the population

Solution

a) The sample is the people who responded to the request.

b) The intended population is likely all persons over 18. The actual population is actually all persons over 18 who have opportunity to receive the request. If the request went out over the internet, for example, the population would be all internet users over 18. If the request went out over several radio and/or TV stations, the population would be all persons over 18 who tune in to those stations.

c) No, since the sample is self-selected it is not likely to be representative of the population. Only those with special interest in abortion issues are likely to spend the time and money to respond to the survey.

Exercise

In the Born Loser cartoon strip by Art Sansom, Brutus expresses joy over an increase in temperature from 1° to 2°. When asked what is so good about 2°, he answers that "it's twice as warm as this morning." explain why Brutus is wrong yet again.

Solution

Temperature ratios are not meaningful because a temperature of 0° does not represent the absence of temperature in the same sense that $0 represents the absence of money. The zero temperature in the exercise (whether Fahrenheit or Centigrade) was determined by a criterion other than "the absence of temperature."

Exercise

A group of students develops a scale for rating the quality of cafeteria food, with 0 representing "neutral: not good and not bad." Bad meals are given negative numbers and good meals are given positive numbers, with the magnitude of the number corresponding to the severity of badness or goodness. The first three meals are rated as 2, 4, and –5. What is the level of measurement for such rating? Explain your choice.

Solution

This is example of ordinal data. It is not interval data because differences are not meaningful. The difference between the ratings +4 and +5 does not necessarily represent the same differential in the quality of food as the difference between 0 and +1.

Exercise

Suppose that a study based on a sample from a targeted population shows that people who own a fax machine have more money than people who do not

a) Make an inference based on the results of this study?

b) What might this inference incorrectly imply?

Solution

a) People who own a fax machine have more money than people who do not

b) This inference may incorrectly inply that if you owned a fax machine, you would have more money than if you did not.

Exercise

Determine whether the statement is true or false, rewrite it as a true statement

a) Data at the ordinal level are quantitative only

b) More types of calculations can be performed with data at the nominal level than with data at the interval level

Solution

a) False. Data at the ordinal level can be qualitative or quantitative

b) False. More types of calculations can be performed with data at the interval level than with data at the nominal level.

Exercise

The region of a country with the highest per capita income for the past six years is shown below

Northeast Southern Southwest Southeast Northern Western

a) Determine whether the data are qualitative or quantitative and identify the data set's level of measurement

b) What is the data set's level of measurement?

Solution

a) Qualitative

b) Nominal

Exercise

The region of a country with the six highest level of food production last year are shown below

1. Eastern 2. Southwest 3. Western 4. Southeast 5. Northwest 6. Southern

a) Determine whether the data are qualitative or quantitative and identify the data set's level of measurement

b) What is the data set's level of measurement?

Solution

a) Quantitative
b) Ordinal

Exercise

The region of a country with the six highest level of food production last year are shown below

22.8 26.4 24.1 22.2 21.6 21.1 25.8 21.5 24.6

a) Determine whether the data are qualitative or quantitative and identify the data set's level of measurement

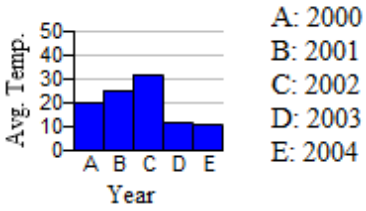
b) What is the data set's level of measurement?

Solution

a) Quantitative
b) Ratio

Exercise

The graph shows the average temperature in an artic city, in degree Fahrenheit, for certain years. Identify the level of measurement of the data listed on the horizontal axis in the graph



Solution

Interval

Exercise

Identify the level of measurement of the data:

a) Temperature
b) Age
c) Family history of illness
d) Pain level (scale of 0 to 10)

Solution

a) Interval
b) Ratio
c) Nominal
d) Ordinal

Exercise

A study was conducted in which 20,211 18-years old male military were given an exam to measure IQ. In addition, the recruits were asked to disclose their smoking status. An individual was considered a smoker if he smoked at least once cigarette per day. The goal of the study was to determine whether adolescents aged 18 to 21 who smoked have a lower IQ than nonsmokers. It was found that the average IQ of the smokers was 94, while he average IQ of the nonsmokers was 101. The researchers concluded that lower IQ individuals are more likely to choose to smoke, not that smoking makes people less intelligent.

a) What is the research objective?
b) What is the population being studied? What is the sample?
c) What are the descriptive statistics?
d) What are the conclusions of the study?

Solution

a) The research objective is to determine if adolescents who smoke have a lower IQ than nonsmokers.
b) The population is all adolescents aged 18-21. The sample consisted of 20,211 18 years old military recruits.
c) Descriptive statistics: The average IQ of the smokers was 94, and the average IQ of nonsmokers was 101.
d) The conclusion is that individuals with a lower IQ are more likely to choose to smoke.

Exercise

Determine whether the variable is qualitative, continuous, or discrete. The following represent information on smart phones.

Model	***Weight*** (oz.)	***Service Provider***	***Depth*** (in)
Motorola Droid *X*	5.47	Verizon	0.39
Motorola Droid 2	5.96	Verizon	0.53
Apple iPhone 4	4.8	ATT	0.37
Samsung Epic 4G	5.5	Sprint	0.6
Samsung Captivate	4.5	ATT	0.39

Solution

Individuals: Model (Motorola Droid X & 2, Apple, and Samsung)
Variables: Weight (oz), Service Provider, Depth (in).
Data for weight: 5.47, 5.96, 4.8, 5.5, 4.5 (oz.)
Data for service provider: Verizon, Verizon, ATT, sprint, ATT
Data for depth: 0.39, 0.53, 0.37, 0.6, 0.39 (in)

The variable weight is ***continuous***.
The variable service provider is ***qualitative***.
The variable depth is ***continuous***.