

Chapter 1: The language of statistics

Multiple choice

1. A single value collected in the context of research is a _____, whereas multiple values collected in the context of research are _____ .

- a. parameter, statistic
- b. data, datum
- c. statistic, parameter
- d. datum, data**

<Remember>

2. A collection of data with different values based upon their source is a _____ .

- a. statistic
- b. variable**
- c. population
- d. parameter

<Remember>

3. Which of the following would be considered quantitative data?

- a. 1st, 3rd, 2nd, 4th
- b. yes, no, no, yes
- c. 5000, 7000, 9000, 3000**
- d. all of these

<Understand>

4. Which of the following would be considered quantitative data?

- a. 1st, 3rd, 2nd, 4th
- b. yes, no, no, yes
- c. 5000, 7000, 9000, 3000**
- d. all of these

<Understand>

5. Which of the following would be considered qualitative data?

a. 0.1, 2.5, 3.6, 0.8

b. blue, green, red, purple

c. 34, 8, 15, 12

d. all of these

<Understand>

6. Which of the following could be a variable?

a. height

b. annual salary

c. attitude

d. all of these

<Apply>

7. The four scales of measurement, from least specific to most specific, are _____.

a. ratio, interval, ordinal, nominal

b. ordinal, nominal, interval, ratio

c. nominal, ordinal, interval, ratio

d. ratio, interval, nominal, ordinal

<Remember>

8. Which of the following would most likely be considered nominal data?

a. age

b. height

c. salary

d. gender

<Understand>

9. Nominal data would be considered dichotomous when _____.

- a. labels have an order
- b. the data are quantitative
- c. the data have only two possible values**
- d. the data are ranked

<Remember>

10. Which of the following would be considered ordinal data (but not interval data)?

- a. ranked product preferences**
- b. favourite colour
- c. sales in euros
- d. all of these

<Apply>

11. Interval data _____.

- a. are quantitative
- b. have meaningful distances between values
- c. have all the properties of ordinal data
- d. all of these**

<Remember>

12. The difference between interval and ratio data is _____.

- a. ratio data have a meaningful zero whereas interval data do not**
- b. ratio data are qualitative whereas interval data are quantitative
- c. ratio data are dichotomous whereas interval data are not
- d. There is no difference between interval and ratio data.

<Understand>

13. If “10” in a variable is twice as many as “5” in that variable, what is that variable’s scale of measurement?

- a. interval
- b. ratio**

c. nominal

d. ordinal

<Apply>

14. A common type of numerically-scored survey item is a _____.

a. ratio scale

b. Likert-type scale

c. Pearson-type scale

d. discrete scale

<Remember>

15. Which of the following would be an example of continuous data?

a. degrees Celsius

b. number of applicants

c. favourite colour

d. gender

<Understand>

16. Data that cannot be divided beyond a certain point would be considered _____.

a. discrete

b. continuous

c. quantitative

d. dichotomous

<Understand>

17. A _____ is a theoretical group that you want to draw conclusions about.

a. sample

b. subject

c. population

d. dataset

<Remember>

18. Why do we examine samples rather than populations?

- a. because populations cannot generally be measured**
- b. because populations provide only qualitative data
- c. because samples are more accurate
- d. because samples include all subjects of interest

<Understand>

19. A construct is _____.

- a. a characteristic or property of interest in a population**
- b. a measurement of how well characteristics of a sample represent characteristics of a population
- c. a group gathered at random from a population
- d. none of these

<Remember>

20. A relationship between constructs can be considered a _____.

- a. operational definition
- b. theory**
- c. hypothesis
- d. experiment

<Remember>

21. Which of the following could be considered an operational definition for customer satisfaction?

- a. a survey item asking about satisfaction on a 5-point scale
- b. number of customer complaints
- c. number of returns
- d. all of these**

<Apply>

22. A hypothesis _____.

a. tests the relationship between operational definitions

b. tests the relationship between constructs

c. measures how well a sample reflects a population

d. defines characteristics of interest in a population

<Remember>

23. Which of the following would be considered an experiment?

a. measuring the relationship between gender and salary

b. measuring the relationship between annual sales and position

c. measuring the difference in attitudes between customers randomly chosen to receive a free sample

d. measuring the difference in attitudes between men and women

<Understand>

24. If you randomly split a sample of eight people into four groups, how many conditions do you have?

a. 0

b. 4

c. 8

d. 32

<Understand>

25. The only approach that allows us to determine cause and effect is a _____.

a. quasi-experiment

b. experiment

c. correlational study

d. longitudinal study

<Remember>

26. If you randomly assign subjects to different groups, you are _____.

a. placing subjects into experimental conditions

- b. selecting subjects from a population
- c. creating associations for a correlational study
- d. none of these

<Understand>

27. If you were running an experiment in which employees completed a training program or no training program at all, participants who completed the training program would be in the _____ condition.

- a. correlational
- b. control

c. treatment

- d. operational

<Understand>

28. If you were create study where customers were randomly assigned to receive a free sample or not receive a free sample, those that did not receive a free sample would be in the _____ condition.

- a. correlational

b. control

- c. treatment
- d. operational

<Understand>

29. If you collected data on the age of your customers and how much they enjoyed shopping at your business, this would be an example of a(n) _____.

- a. condition
- b. experiment
- c. quasi-experiment

d. correlational study

<Understand>

30. The difference between an experiment and a quasi-experiment is that _____.
- a. you can make conclusions about causality in a quasi-experiment, but not an experiment
 - b. you use more than one condition in a quasi-experiment, but not in an experiment
 - c. you can assess more than one variable in an experiment, but not in a quasi-experiment
 - d. you use random assignment in an experiment, but not in a quasi-experiment**
- <Understand>

True/False

- 1. A dataset is a collection of data linked together in a meaningful way.
 - a. True**
 - b. False

- 2. A variable is a collection of data linked together in a meaningful way.
 - a. True
 - b. False**

- 3. A case is a source of data about one or more variables.
 - a. True**
 - b. False

- 4. Qualitative data refer to qualities, like letters or words.
 - a. True**
 - b. False

- 5. There are four scales of measurement.
 - a. True**
 - b. False

- 6. Favourite colour is an example of a dichotomous variable.
 - a. True

b. False

7. Ratio data has all of the qualities of ordinal data but not interval data.

a. True

b. False

8. Degrees Celsius are an example of a ratio-level scale of measurement.

a. True

b. False

9. Data that can be divided infinitely while remaining meaningful are continuous data.

a. True

b. False

10. Ten subjects drawn at random from a larger population would be considered a sample.

a. True

b. False

11. A representative sample would be one that accurately reflects the characteristics of a population.

a. True

b. False

12. A statement specifying the relationship between constructs would be considered a hypothesis.

a. True

b. False

13. The following statement is an example of a theory: "There is a relationship between happiness and job success."

a. True

b. False

14. If you operationalize a construct, you are defining it in a “real world” way.

a. True

b. False

15. The following statement is an example of a hypothesis: “Employees who report being happier on a 5-point Likert-type scale will also report being more satisfied on a 5-point Likert-type scale.”

a. True

b. False

16. You should not draw conclusions about causality from an experiment.

a. True

b. False

17. If you randomly assign participants to three conditions, you are conducting an experiment.

a. True

b. False

18. If you collect data from participants on more than one variable without making any changes, you are conducting a correlational study.

a. True

b. False

19. A quasi-experiment uses random assignment.

a. True

b. False

20. Quasi-experiment should always be your first choice for study design.

a. True

b. False

Short Answer/Essay

1. Define and give an example of each of the four scales of measurement.

Nominal data have meaningful labels. Example: gender (male/female).

Ordinal data are nominal data with a meaningful order. Example: rank ordering of sales performance (1st, 2nd, 3rd)

Interval data are ordinal data with meaningful distances between numbers. Example: the result from a Likert-type survey (with ratings 1–5)

Ratio data are interval data with a meaningful zero. Example: sales performance in British pounds

<Remember> <Apply>

2. Some statisticians consider a Likert-type item to be an ordinal-level scale, whereas others consider it to be an interval-level scale. Why?

Interval scales require equal distances between values whereas ordinal scales only require meaningful order. People in favour of “interval” argue that the distances between numbers in a Likert-type survey item (for example, 1 through 5) are equidistant; that is, the distance between 1 and 2 is the same as the distance between 4 and 5. People in favour of “ordinal” argue that this isn’t true; all we know is that “4” is more than “3”, not how far apart they are. However, if these distances are treated as equal, we are able to conduct more powerful statistical tests on the data.

<Understand>

3. Why is it important that samples are representative from the population? Define both sample and population and provide an example of each.

Populations are groups of interest that we wish to make conclusions about. Samples are groups chosen at random from the population. We want samples to be representative of the population so that we can make accurate conclusions about the population using the sample. An example of a population would be “all customers of our store” whereas an example of a sample would be “customers that chose to complete our survey.”

<Understand> <Apply>

4. Explain the difference between an experiment, correlational study, and a quasi-experiment. Give an example of each that could be used to address the following research question: “Does training impact job satisfaction?”

Correlational studies examine pre-existing relationships. Quasi-experiments compare pre-existing groups. Experiments compare groups randomly assigned to different conditions. A correlational training study might examine the relationship between number of training sessions completed and job satisfaction as measured on a survey. A quasi-experiment might compare job satisfaction between those that chose to complete training and those that chose not to. An experiment might compare job satisfaction between those randomly assigned to complete training (treatment group) and those that were randomly assigned not to (control group).

<Remember> <Apply>

Data

For the following questions, consider the table below:

	Training	Test score
1	A	15
2	A	14
3	B	19
4	B	20
5	C	10
6	C	11

1. The table above, as a whole, depicts a _____.
 - a. data
 - b. datum

c. dataset

d. variable

2. In the table above, “19” is a _____.

a. data

b. datum

c. dataset

d. variable

3. In the table above, what is “Training”?

a. data

b. datum

c. dataset

d. variable

4. What is the value of “Training” for Case 5?

a. 10

b. B

c. 20

d. C

5. Which of the following describes the numbers under the “Training” heading?

a. qualitative

b. ordinal

c. interval

d. quantitative

6. If employees were randomly assigned to Training A, B or C, this table depicts _____.

a. a correlational study

b. a quasi-experiment

c. an experiment

d. none of these

7. If employees chose whether they wished to complete Training A, B or C, this table depicts

_____.

a. a correlational study

b. a quasi-experiment

c. an experiment

d. none of these

8. How many cases are depicted in this table?

a. 2

b. 6

c. 12

d. 18

9. How many variables are depicted in this table?

a. 2

b. 6

c. 12

d. 18

10. How many data are depicted in this table?

a. 2

b. 6

c. 12

d. 18

Chapter 2: Working with numbers and data display

Multiple choice

1. Why is it important to accurately depict data?
- a. because you can be misled by others' poor translations of that data
 - b. because you can mislead others with your poor translations of that data
 - c. because you could mislead yourself with your poor translations or that data

d. all of these

<Understand>

2. A concise definition of a statistical concept using statistical notation is a _____.

a. formula

- b. operational definition
- c. graph
- d. none of these

<Remember>

3. The appropriate way to represent "three-quarters" as a fraction is _____.

a. 0.75

b. $\frac{3}{4}$

c. 75%

d. three-quarters

<Remember>

4. To convert a proportion to a percentage, you should _____.

a. do nothing – they are the same.

b. divide it by 100 and add a "%"

c. multiply it by 100 and add a "%"

d. divide the number in a calculator and report the resulting value

<Remember>

5. Which of the following represents population size?

- a. n
- b. N**
- c. p
- d. P

<Remember>

6. If $n = 100$, what can you conclude?

- a. There are 100 subjects in the population.
- b. There are 100 samples in the dataset.
- c. There are 100 subjects in the sample.**
- d. There are 100 variables in the dataset.

<Understand>

7. A frequency is _____.

- a. a portion of a whole, represented as a decimal
- b. a count of how many times a value appears in a variable**
- c. a table containing all categories
- d. an illustration of qualitative data

<Remember>

8. Which of the following symbols represents a frequency?

- a. f**
- b. p
- c. n
- d. N

<Remember>

9. For the following dataset, what does f equal for the value “3”?

[2, 4, 1, 2, 5, 3, 4, 3, 1, 4]

- a. 1

b. 2

c. 0.1

d. 0.2

<Apply>

10. For the following dataset, what does $rel. f$ equal for the value “5”?

[2, 3, 5, 1, 2, 3, 5, 1, 2, 5]

a. 2

b. 3

c. 0.2

d. 0.3

<Apply>

11. If you want to convert from a frequency to a relative frequency, you should _____.

a. count how many times a value appears in your dataset

b. divide the frequency by the sample size

c. divide the sample size by the frequency

d. count the total number of values in the dataset

<Remember>

12. For the following dataset, what does $cum. f$ equal for the value “5”?

[3, 1, 4, 5, 2, 1, 3, 5, 2, 1]

a. 2

b. 0.2

c. 0.8

d. 1

<Apply>

13. If you want to make a frequency table, what is the first thing you should do?

a. determine the frequencies of each of your values

b. calculate the relative frequencies of each of your values

c. make a list of every value and order them

d. calculate the cumulative frequencies of all of your values

<Remember>

14. How many values should be included in a frequency table for the following data?

[1st, 5th, 5th, 2nd, 4th, 2nd, 5th, 1st, 2nd]

a. 4

b. 5

c. 6

d. all of these

<Apply>

15. In a bar chart, what is represented on the x -axis?

a. frequencies

b. relative frequencies

c. variable categories

d. proportions

<Remember>

16. When should you translate data with a bar chart?

a. when you have qualitative data

b. when you have nominal data

c. when you have ordinal data

d. all of these

<Understand>

17. To ensure that your bar chart is easily interpretable, your bar chart should have which of the following attributes?

a. It should be wider than it is tall.

b. The axis labels should be left-aligned.

c. The largest value should be at the bottom of the chart.

d. Each bar should be a different colour.

<Remember>

18. An illustration of qualitative data representing a variable's categories as portions of a circle is a _____.

a. bar chart

b. histogram

c. frequency polygon

d. pie chart

<Remember>

19. To ensure that your pie chart is easily interpretable, your pie chart should have which of the following attributes?

a. The chart should contain a legend.

b. Each wedge should be the same colour as every other wedge.

c. Each wedge should contain only a relative frequency.

d. The pie should be taller than it is wide.

<Remember>

20. If you want to illustrate a quantitative variable, you should use a _____.

a. bar chart

b. pie chart

c. frequency table

d. frequency polygon

<Remember>

21. The primary difference between a histogram and a bar chart is that _____.

a. there is no gap between the bars on a histogram.

b. histograms have an x - and y -axis.

c. the y -axis represents frequencies on a histogram.

d. the histogram is used for qualitative data.

<Remember>

22. For which of the following variables would a histogram be appropriate?

- a. favourite colour
- b. job title
- c. month
- d. age in years**

<Apply>

23. The range of values represented by a single bar in a histogram is called a _____.

- a. distribution
- b. frequency
- c. bin**
- d. box

<Remember>

24. Frequency polygons are similar to histograms except _____.

- a. frequency polygons use lines instead of bars to represent frequencies**
- b. histograms can be used represent more than two frequency distributions
- c. frequency polygons represent frequencies on the x -axis
- d. frequency polygons represent data more clearly than a histogram

<Understand>

25. If you want to illustrate the relationship between variables, you should use a _____.

- a. bar chart
- b. pie chart
- c. histogram
- d. scatterplot**

<Remember>

26. Each point in a scatterplot represents _____.

a. the frequency of a variable

b. the scores of two variables for a single case

c. the scores of two cases for a single variable

d. the proportions of scores of two cases for a single variable

<Remember>

27. Which of these should you use to illustrate responses to the following question? “How many sales were made last year?”

a. bar chart

b. pie chart

c. histogram

d. scatterplot

<Apply>

28. Which of these should you use to illustrate responses to the following question? “How many employees are working at each branch of our company?”

a. scatterplot

b. frequency polygon

c. bar chart

d. histogram

<Apply>

29. Which of these should you use to address the following question? “What is the relationship between number of sales and employee job satisfaction?”

a. scatterplot

b. bar chart

c. pie chart

d. frequency polygon

<Apply>

30. Which of these should you use to address the following question? “How does the distribution of salaries last year compare with the distribution of salaries this year?”

- a. scatterplot
- b. frequency polygon**
- c. pie chart
- d. bar chart

<Apply>

True/False

1. In order to help interpret large amounts of raw data, it is helpful to translate data into illustrations, pictures, or graphics.

- a. True**
- b. False

2. Quantitative and qualitative variables are treated the same when illustrated graphically.

- a. True
- b. False**

3. When writing formulas, it is more common to use proportions than it is to use fractions.

- a. True**
- b. False

4. 5% is another representation of the following proportion: 0.5

- a. True
- b. False**

5. If $N = 500$, there are 500 subjects in the sample.

- a. True
- b. False**

6. The number of cases in a sample is the sample size.

a. True

b. False

7. If $n = 20$, and 5 appears 4 times in the variable being examined, the rel. f of 5 for that variable is .25.

a. True

b. False

8. If $n = 20$ and $f = 4$ for the value 2, the relative frequency of 2 is 5.

a. True

b. False

9. A count of how many times a value and all values below it appears in a variable is the cumulative frequency.

a. True

b. False

10. A frequency table consists of only simple frequencies.

a. True

b. False

11. The most common illustration for qualitative data is a pie chart.

a. True

b. False

12. When creating a bar chart, your bars should touch.

a. True

b. False

13. One of the most common ways to mislead people with graphics is to use a value other than 0 as the smallest value on the y -axis.

a. True

b. False

14. You should use similar colours on your pie chart so that the chart is more visually appealing.

a. True

b. False

15. If you have quantitative data, you should use a histogram instead of a bar chart.

a. True

b. False

16. The y-axis for a histogram should include counts or frequencies.

a. True

b. False

17. Histograms should always use a bin size of 10.

a. True

b. False

18. If you want to compare two or more frequency distributions, you should use a frequency polygon instead of a histogram.

a. True

b. False

19. If you want to know what happens to the y-variable as the x-variable increases, you should translate your data with a pie chart.

a. True

b. False

20. Scatterplots can be used with qualitative or quantitative data.

a. True

b. False

Short Answer/Essay

1. Your friend wants to know why he should bother learning about statistics. Can't he just hire someone else to do that for him? Explain why your friend would benefit by learning about appropriate data translation and provide an example of how this knowledge could help him.

Statistics, including figures and tables, can be used to mislead others. Without learning statistics, my friend could be misled. For example, a graph might be used that does not clearly label the values on its axes to make differences appear larger than they really are.

<Apply>

2. List the steps in creating a frequency table. Explain how to calculate values at each step.

1. Column 1: List all the possible values for the variable in ascending order.

2. Column 2: Count the number of times each value you listed in Column 1 appears in the original dataset and record these values under f .

3. Column 3: Divide the f values you just calculated by n to determine rel. f .

4. Column 4: Count the number of values for each row and all the rows above it to determine cum. F .

<Remember>

3. Explain the major differences between a bar chart and a histogram, including when to use each.

There are two primary differences. First, visually, the bars touch (there are no gaps between bars). Second, the bars in a bar chart represent discrete values whereas the bars in a histogram are really bins (ranges of values). We use bar charts for qualitative variables and histograms for quantitative variables.

<Understand>

4. Explain the difference between a frequency polygon and a scatterplot, including when to use each.

A frequency polygon is used to summarize the distribution of a single quantitative variable. You can also layer multiple lines on top of each other to compare the distributions to each other. A scatterplot is used to compare two variables directly to one another, with each point in the plot representing a single case – they can also be used with a mix of qualitative and quantitative variables. Frequency polygons are most useful when you want to compare two quantitative distributions. Scatterplots are most useful when you want to compare two variables where every case has values for both variables.

<Understand>

Data

For the following questions, consider the dataset below collected from a sample of trainees:

	Training	Test score
1	A	12
2	A	9
3	B	6
4	B	10
5	C	30

Test scores represent the number of test questions trainees answered correctly on the post-training survey.

1. The scale of measurement for the Training variable above is _____.

- a. qualitative
- b. quantitative
- c. nominal**
- d. ratio

2. The best translation for the Training variable above would be _____.

- a. a bar chart**

b. a histogram

c. a scatterplot

d. a frequency polygon

3. The scale of measurement for the Test Score variable above is _____.

a. qualitative

b. quantitative

c. nominal

d. ratio

4. The best translation for the Test Score variable above would be _____.

a. a histogram

b. a bar chart

c. a scatterplot

d. a pie chart

5. The best way to compare the Training variable and the Test Score variable would be _____.

a. a bar chart

b. a histogram

c. a scatterplot

d. a frequency polygon

6. What is the frequency of Training C?

a. 0.1

b. 0.2

c. 1

d. 2

7. What is the relative frequency of Training C?

a. 0.1

b. 0.2

- c. 1
- d. 2

8. What is the cumulative frequency of Training C?

- a. 0.1
- b. 0.2
- c. 1**
- d. 2

9. What is n in the dataset above?

- a. 2
- b. 5**
- c. 10
- d. cannot be determined from the dataset provided

10. What is N in the dataset above?

- a. 2
- b. 5
- c. 10
- d. cannot be determined from the dataset provided**