Introductory Statistics 2nd edition Illowsky Solutions Manual

SKU: 9781711472577-SOLUTIONS

openstax™

Introductory Statistics 2e.

Introductory Statistics 2nd edition Illowsky Solutions Manual

CHAPTER 1: SAMPLING AND DATA

Exercise 1. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term "population" refers to in the example for Researcher A.

Solution patients with the virus

Exercise 2. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term "sample" refers to in the example for Researcher A.

Solution patients with the virus sampled from researcher A and researcher B.

Exercise 3. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40*

Introductory Statistics 2nd edition Illowsky Solutions Manual

patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term "parameter" refers to in the example for Researcher A.

Solution — the average length of time (in days) patients with the virus live after treatment

Exercise 4. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term "statistic" refers to in the example for Researcher A.

Solution — the average length of time (in days) patients with the virus from the sample live after treatment

Exercise 5. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*

Introductory Statistics 2nd edition Illowsky Solutions Manual

Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term "variable" refers to in the example for Researcher A.

Solution X = the length of time (in days) patients with the virus live after treatment

Exercise 6. *"Number of times per week" is what type of data?*
a. qualitative; b. quantitative discrete; c. quantitative continuous

Solution b

Exercise 7. *A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Antonio, Texas. The first house in the neighborhood around the park was selected randomly, and then the resident of every eighth house in the neighborhood around the park was interviewed. The sampling method was*
a. simple random; b. systematic; c. stratified; d. cluster

Solution b

Exercise 8. *A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Antonio, Texas. The first house in the neighborhood around the park was selected randomly, and then the resident of every eighth house in the neighborhood around the park was interviewed. "Duration (amount of time)" is what type of data?*
a. qualitative; b. quantitative discrete; c. quantitative continuous

Solution c

Exercise 9. *A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Antonio, Texas. The first house in the neighborhood around the park was selected randomly, and then the resident of every eighth house in the neighborhood around the park was interviewed. The colors of the houses around the park are what kind of data?*
a. qualitative; b. quantitative discrete; c. quantitative continuous

Solution A

Exercise 10. *A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Antonio, Texas. The first house in the neighborhood around the park was selected*

Introductory Statistics 2nd edition Illowsky Solutions Manual

*randomly, and then the resident of every eighth house in the neighborhood around the park was interviewed. The population is*_____

Solution the houses in the neighborhood around the park in San Antonio, Texas.

Exercise 11. *Table 1.27 contains the total number of deaths worldwide as a result of earthquakes over a 13-year period.*

Year	Total Number of Deaths
1	231
2	21,357
3	11,685
4	33,819
5	228,802
6	88,003
7	6,605
8	712
9	88,011
10	1,790
11	320,120
12	21,953
13	768
Total	**823,856**

Table 1.27

a. What is the proportion of deaths between Year 8 and Year 13?
b. What percent of deaths occurred before Year 2?
c. What is the percent of deaths that occurred in Year 4 or after Year 11?
d. What is the fraction of deaths that happened before Year 13?
e. What kind of data is the number of deaths?
f. Earthquakes are quantified according to the amount of energy they produce (examples are 2.1, 5.0, 6.7). What type of data is that?
g. What contributed to the large number of deaths in Year 11? In Year 5? Explain.

Solution
a. 0.5242
b. 0.03%
c. 6.86%
d. $\frac{823,088}{823,856}$
e. quantitative discrete
f. quantitative continuous
g. In both years, underwater earthquakes produced massive tsunamis.

Introductory Statistics 2nd edition Illowsky Solutions Manual

Exercise 12. *Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience). A group of test subjects is divided into twelve groups; then four of the groups are chosen at random.*

Solution cluster

Exercise 13. *Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience). A market researcher polls every tenth person who walks into a store.*

Solution systematic

Exercise 14. *Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience). The first 50 people who walk into a sporting event are polled on their television preferences.*

Solution Convenience

Exercise 15. *Determine the type of sampling used (simple random, stratified, systematic, cluster, or convenience). A computer generates 100 random numbers, and 100 people whose names correspond with the numbers on the list are chosen.*

Solution simple random

Exercise 16. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Complete the tables using the data provided:

Introductory Statistics 2nd edition Illowsky Solutions Manual

Survival Length (in days)	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0.5–6.5			
6.5–12.5			
12.5–18.5			
18.5–24.5			
24.5–30.5			
30.5–36.5			
36.5–42.5			
42.5–48.5			

Table 1.28 Researcher A

Survival Length (in days)	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0.5–6.5			
6.5–12.5			
12.5–18.5			
18.5–24.5			
24.5–30.5			
30.5–36.5			
36.5–45.5			

Table 1.29 Researcher B

Solution

Survival Length (in days)	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0.5–6.5	2	$\frac{2}{40}$	0.0500
6.5–12.5	5	$\frac{5}{40}$	0.1750
12.5–18.5	9	$\frac{9}{40}$	0.4000
18.5–24.5	5	$\frac{5}{40}$	0.5250
24.5–30.5	7	$\frac{7}{40}$	0.7000
30.5–36.5	7	$\frac{7}{40}$	0.8750
36.5–42.5	2	$\frac{2}{40}$	0.9250

Introductory Statistics 2nd edition Illowsky Solutions Manual

Survival Length (in days)	Frequency	Relative Frequency	Cumulative Relative Frequency
42.5–48.5	3	$\frac{3}{40}$	1.0000
0.5–6.5	3	$\frac{3}{40}$	0.0750
6.5–12.5	2	$\frac{2}{40}$	0.1250
12.5–18.5	11	$\frac{11}{40}$	0.4000
18.5–24.5	8	$\frac{8}{40}$	0.6000
24.5–30.5	7	$\frac{7}{40}$	0.7750
30.5–36.5	5	$\frac{5}{40}$	0.9000
36.5–45.5	4	$\frac{4}{40}$	1

Exercise 17. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Determine what the key term data refers to in the above example for Researcher A.

Solution values for X, such as 3, 4, 11, and so on

Exercise 18. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the*

Introductory Statistics 2nd edition Illowsky Solutions Manual

symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
List two reasons why the data may differ.

Solution

Answers will vary. Sample answer: One reason may be the average age of the individuals in the two samples. Or, perhaps the drug affects men and women differently. If the ratio of men and women aren't the same in both sample groups, then the data would differ.

Exercise 19.

Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Can you tell if one researcher is correct and the other one is incorrect? Why?

Solution

No, we do not have enough information to make such a claim.

Exercise 20.

Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*

Introductory Statistics 2nd edition Illowsky Solutions Manual

Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Would you expect the data to be identical? Why or why not?

Solution — Since the treatment is not the same the data might be different unless neither treatment has an effect.

Exercise 21. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
How might the researchers gather random data?

Solution — Take a simple random sample from each group. One way is by assigning a number to each patient and using a random number generator to randomly select patients.

Exercise 22. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Suppose that the first researcher conducted his survey by randomly choosing one state in the nation and then randomly picking 40 patients from that state. What sampling method would that researcher have used?

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution He has used a simple random sample method.

Exercise 23. *Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new drug is currently under study to address a respiratory virus. It is given to patients once the patient exhibits symptoms of the virus. Of interest is the average (mean) length of time in days from the time the patient starts the treatment until the symptoms are alleviated. Two researchers each follow a different set of 40 patients with the respiratory virus from the start of treatment until the symptoms are alleviated. The following data (in days) are collected.*
Researcher A: *3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34*
Researcher B: *3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29*
Suppose that the second researcher conducted his survey by choosing 40 patients he knew. What sampling method would that researcher have used? What concerns would you have about this data set, based upon the data collection method?

Solution This would be convenience sampling and is not random.

Exercise 24. *Two researchers are gathering data on hours of video games played by teens. They each randomly sample different groups of 150 students from the same high school. They collect the following data.*

Hours Played per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.30 Researcher A

Hours Played per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0–2	48	0.32	0.32
2–4	51	0.34	0.66
4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97

Introductory Statistics 2nd edition Illowsky Solutions Manual

10–12	4	0.03	1

Table 1.31 Researcher B

Give a reason why the data may differ.

Solution The researchers are studying different groups, so there will be some variation in the data.

Exercise 25. *Two researchers are gathering data on hours of video games played by teens. They each randomly sample different groups of 150 students from the same high school. They collect the following data.*

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.30 Researcher A

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	48	0.32	0.32
2–4	51	0.34	0.66
4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97
10–12	4	0.03	1

Table 1.31 Researcher B

Would the sample size be large enough if the population is the students in the school?

Solution Yes, the sample size of 150 would be large enough to reflect a population of one school.

Exercise 26. *Two researchers are gathering data on hours of video games played by teens. They each randomly sample different groups of 150 students from the same high school. They collect the following data.*

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**

Introductory Statistics 2nd edition Illowsky Solutions Manual

0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.30 Researcher A

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	48	0.32	0.32
2–4	51	0.34	0.66
4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97
10–12	4	0.03	1

Table 1.31 Researcher B

Would the sample size be large enough if the population is teens in the United States?

Solution There are many school-aged children and young adults in the United States, and the study was done at only one school, so the sample size is not large enough to reflect the population.

Exercise 27. *Two researchers are gathering data on hours of video games played by teens. They each randomly sample different groups of 150 students from the same high school. They collect the following data.*

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.30 Researcher A

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	48	0.32	0.32
2–4	51	0.34	0.66

Introductory Statistics 2nd edition Illowsky Solutions Manual

4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97
10–12	4	0.03	1

Table 1.31 Researcher B

Researcher A concludes that most teens play video games between four and six hours each week. Researcher B concludes that most teens play video games between two and four hours each week. Who is correct?

Solution

Even though the specific data support each researcher's conclusions, the different results suggest that more data need to be collected before the researchers can reach a conclusion.

Exercise 28.

Two researchers are gathering data on hours of video games played by teens. They each randomly sample different groups of 150 students from the same high school. They collect the following data.

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	26	0.17	0.17
2–4	30	0.20	0.37
4–6	49	0.33	0.70
6–8	25	0.17	0.87
8–10	12	0.08	0.95
10–12	8	0.05	1

Table 1.30 Researcher A

Hours Played per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
0–2	48	0.32	0.32
2–4	51	0.34	0.66
4–6	24	0.16	0.82
6–8	12	0.08	0.90
8–10	11	0.07	0.97
10–12	4	0.03	1

Table 1.31 Researcher B

As part of a way to reward students for participating in the survey, the researchers gave each student a gift card to a video game store. Would this affect the data if students knew about the award before the study?

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution

Yes, people who play games more might be more likely to participate, since they would want the gift card more than a student who does not play video games. This would leave out many students who do not play games at all and skew the data.

Exercise 29.

A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke patients recovering over a period of several weeks. The first study collected the data in Table 1.31. The second study collected the data in Table 1.32.

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	142	43	15
Did not use program	72	110	18

Table 1.32

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	105	74	19
Did not use program	89	99	14

Table 1.32

Given what you know, which study is correct?

Solution

There is not enough information given to judge if either one is correct or incorrect.

Exercise 30.

A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke patients recovering over a period of several weeks. The first study collected the data in Table 1.31. The second study collected the data in Table 1.32.

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	142	43	15
Did not use program	72	110	18

Table 1.32

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	105	74	19

Introductory Statistics 2nd edition Illowsky Solutions Manual

Did not use program	89	99	12

Table 1.33

The first study was performed by the company that designed the software program. The second study was performed by the American Medical Association. Which study is more reliable?

Solution

The second study is more reliable, because the company would be interested in showing results that favored a higher rate of improvement from patients using their software. The data may be skewed; however, the American Medical Association is not concerned with the success of the software and so should be objective.

Exercise 31.

A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke patients recovering over a period of several weeks. The first study collected the data in Table 1.31. The second study collected the data in Table 1.32.

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	142	43	15
Did not use program	72	110	18

Table 1.32

Group	**Showed improvement**	**No improvement**	**Deterioration**
Used program	105	74	19
Did not use program	89	99	12

Table 1.33

Both groups that performed the study concluded that the software works. Is this accurate?

Solution

The software program seems to work because the second study shows that more patients improve while using the software than not. Even though the difference is not as large as that in the first study, the results from the second study are likely more reliable and still show improvement.

Exercise 32.

A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke

Introductory Statistics 2nd edition Illowsky Solutions Manual

patients recovering over a period of several weeks. The first study collected the data in Table 1.31. The second study collected the data in Table 1.32.

Group	Showed improvement	No improvement	Deterioration
Used program	142	43	15
Did not use program	72	110	18

Table 1.32

Group	Showed improvement	No improvement	Deterioration
Used program	105	74	19
Did not use program	89	99	12

Table 1.33

The company takes the two studies as proof that their software causes mental improvement in stroke patients. Is this a fair statement?

Solution

No, the data suggest the two are correlated, but more studies need to be done to prove that using the software causes improvement in stroke patients.

Exercise 33.

A pair of studies was performed to measure the effectiveness of a new software program designed to help stroke patients regain their problem-solving skills. Patients were asked to use the software program twice a day, once in the morning and once in the evening. The studies observed 200 stroke patients recovering over a period of several weeks. The first study collected the data in Table 1.31. The second study collected the data in Table 1.32.

Group	Showed improvement	No improvement	Deterioration
Used program	142	43	15
Did not use program	72	110	18

Table 1.32

Group	Showed improvement	No improvement	Deterioration
Used program	105	74	19
Did not use program	89	99	12

Table 1.33

Patients who used the software were also a part of an exercise program whereas patients who did not use the software were not. Does this change the validity of the conclusions from Exercise 1.31?

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution — Yes, because we cannot tell if the improvement was due to the software or the exercise; the data is confounded, and a reliable conclusion cannot be drawn. New studies should be performed.

Exercise 34. *Is a sample size of 1,000 a reliable measure for a population of 5,000?*

Solution — Yes, 1,000 represents 20% of the population and should be representative, if the population of the sample is chosen at random.

Exercise 35. *Is a sample of 500 volunteers a reliable measure for a population of 2,500?*

Solution — No, even though the sample is large enough, the fact that the sample consists of volunteers makes it a self-selected sample, which is not reliable.

Exercise 36. *A question on a survey reads: "Do you prefer the delicious taste of Brand X or the taste of Brand Y?" Is this a fair question?*

Solution — No, the question is creating undue influence by adding the word "delicious" to describe Brand X. The wording may influence responses.

Exercise 37. *Is a sample size of two representative of a population of five?*

Solution — No, even though the sample is a large portion of the population, two responses are not enough to justify any conclusions. Because the population is so small, it would be better to include everyone in the population to get the most accurate data.

Exercise 38. *Is it possible for two experiments to be well run with similar sample sizes to get different data?*

Solution — Yes, there will most likely be a degree of variation between any two studies, even if they are set up and run the same way. Each study may be affected differently by unknown factors such as location, mood of the subjects, or time of year.

Exercise 39. *What type of measure scale is being used? Nominal, ordinal, interval or ratio.*
a. High school soccer players classified by their athletic ability: Superior, Average, Above average
b. Baking temperatures for various main dishes: 350, 400, 325, 250, 300
c. The colors of crayons in a 24-crayon box
d. Social security numbers
e. Incomes measured in dollars
f. A satisfaction survey of a social website by number: 1 = very satisfied, 2 = somewhat satisfied, 3 = not satisfied
g. Political outlook: extreme left, left-of-center, right-of-center, extreme right
h. Time of day on an analog watch

Introductory Statistics 2nd edition Illowsky Solutions Manual

i. The distance in miles to the closest grocery store
j. The dates 1066, 1492, 1644, 1947, and 1944
k. The heights of 21–65 year-old women
l. Common letter grades: A, B, C, D, and F

Solution

a. ordinal
b. interval
c. nominal
d. nominal
e. ratio
f. ordinal
g. nominal
h. interval
i. ratio
j. interval
k. ratio
l. ordinal

Exercise 40. *Design an experiment. Identify the explanatory and response variables. Describe the population being studied and the experimental units. Explain the treatments that will be used and how they will be assigned to the experimental units.*
Describe how blinding and placebos may be used to counter the power of suggestion.

Solution Answers will vary.

Exercise 41. *Discuss potential violations of the rule requiring informed consent.*
a. People who are incarcerated in a correctional facility are offered good behavior credit in return for participation in a study.
b. A research study is designed to investigate a new children's allergy medication.
c. Participants in a study are told that the new medication being tested is highly promising, but they are not told that only a small portion of participants will receive the new medication. Others will receive placebo treatments and traditional treatments.

Solution

a. Inmates may not feel comfortable refusing participation, or may feel obligated to take advantage of the promised benefits. They may not feel truly free to refuse participation.
b. Parents can provide consent on behalf of their children, but children are not competent to provide consent for themselves.
c. All risks and benefits must be clearly outlined. Study participants must be informed of relevant aspects of the study in order to give appropriate consent.

Introductory Statistics 2nd edition Illowsky Solutions Manual

Exercise 42. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
A fitness center is interested in the mean amount of time a client exercises in the center each week.

Solution
a. The population is all of the clients of the fitness center.
b. A sample of the clients that use the fitness center for a given week.
c. The average amount of time that all clients exercise in one week.
d. The average amount of time that a sample of clients exercises in one week.
e. The amount of time that a client exercises in one week.
f. Examples are: 2 hours, 5 hours, and 7.5 hours

Exercise 43. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
Ski resorts are interested in the mean age that children take their first ski and snowboard lessons. They need this information to plan their ski classes optimally.

Solution
a. all children who take ski or snowboard lessons
b. a group of these children
c. the population mean age of children who take their first snowboard lesson
d. the sample mean age of children who take their first snowboard lesson
e. X = the age of one child who takes their first ski or snowboard lesson
f. values for X, such as 3, 7, and so on

Exercise 44. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
A cardiologist is interested in the mean recovery period of their patients who have had heart attacks.

Solution
a. the cardiologist's patients
b. a group of the cardiologist's patients
c. the mean recovery period of all of the cardiologist's patients
d. the mean recovery period of the group of the cardiologist's patients
e. X = the mean recovery period of one patient
f. values for X, such as 10 days, 14 days, 20 days, and so on

Exercise 45. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
Insurance companies are interested in the mean health costs each year of their clients, so that they can determine the costs of health insurance.

Solution
a. the clients of the insurance companies
b. a group of the clients
c. the mean health costs of the clients
d. the mean health costs of the sample

Introductory Statistics 2nd edition Illowsky Solutions Manual

e. X = the health costs of one client
f. values for X, such as 34, 9, 82, and so on

Exercise 46. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
A politician is interested in the proportion of voters in their district who think he is doing a good job.

Solution
a. all voters in the politician's district
b. a random selection of voters in the politician's district
c. the proportion of voters in this district who think this politician is doing a good job
d. the proportion of voters in this district who think this politician is doing a good job in the sample
e. X = the number of voters in the district who think this politician is doing a good job
f. Yes, he is doing a good job. No, he is not doing a good job.

Exercise 47. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
A marriage counselor is interested in the proportion of clients they counsel who stay married.

Solution
a. all the clients of this counselor
b. a group of clients of this marriage counselor
c. the proportion of all of this counselor's clients who stay married
d. the proportion of the sample of the counselor's clients who stay married
e. X = whether a client stayed married
f. yes, no

Exercise 48. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
Political pollsters may be interested in the proportion of people who will vote for a particular cause.

Solution
a. all voters (in a certain geographic area)
b. a random selection of all the voters
c. the proportion of voters who are interested in this particular cause
d. the proportion of voters who are interested in this particular cause in the sample
e. X = the number of voters who are interested in this particular cause
f. yes, no

Exercise 49. *Identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.*
A marketing company is interested in the proportion of people who will buy a particular product.

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution

a. all people (maybe in a certain geographic area, such as the United States)
b. a group of the people
c. the proportion of all people who will buy the product
d. the proportion of the sample who will buy the product
e. X = whether a person bought the product
f. buy, not buy

Exercise 50.

A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter. What is the population she is interested in?
a. all Lake Tahoe Community College students
b. all Lake Tahoe Community College English students
c. all Lake Tahoe Community College students in her classes
d. all Lake Tahoe Community College math students

Solution

d

Exercise 51.

A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter.
Consider the following:
X = number of days a Lake Tahoe Community College math student is absent
In this case, X is an example of a:
a. variable.
b. population.
c. statistic.
d. data.

Solution

a

Exercise 52.

A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter.
The instructor's sample produces a mean number of days absent of 3.5 days. This value is an example of a:
a. parameter.
b. data.
c. statistic.
d. variable.

Solution

c

Exercise 53.

Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.

Introductory Statistics 2nd edition Illowsky Solutions Manual

number of tickets sold to a concert

Solution quantitative discrete, 150

Exercise 54. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
percent of body fat

Solution quantitative continuous, 19.2%

Exercise 55. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
favorite baseball team

Solution qualitative, Oakland A's

Exercise 56. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
time in line to buy groceries

Solution quantitative continuous, 7.2 minutes

Exercise 57. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
number of students enrolled at Evergreen Valley College

Solution quantitative discrete, 11,234 students

Exercise 58. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
most-watched television show

Solution qualitative, Dancing with the Stars

Exercise 59. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
brand of toothpaste

Solution qualitative, Crest

Introductory Statistics 2nd edition Illowsky Solutions Manual

Exercise 60. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
distance to the closest movie theatre

Solution quantitative continuous, 8.32 miles

Exercise 61. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
age of executives in Fortune 500 companies

Solution quantitative continuous, 47.3 years

Exercise 62. *Identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.*
number of competing computer spreadsheet software packages

Solution quantitative discrete, three

Exercise 63. *A study was done to determine the age, number of times per week, and the duration (amount of time) of resident use of a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every 8th house in the neighborhood around the park was interviewed.*
"Number of times per week" is what type of data?
a. qualitative
b. quantitative discrete
c. quantitative continuous

Solution b

Exercise 64. *A study was done to determine the age, number of times per week, and the duration (amount of time) of resident use of a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every 8th house in the neighborhood around the park was interviewed.*
"Duration (amount of time)" is what type of data?
a. qualitative
b. quantitative discrete
c. quantitative continuous

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution c

Exercise 65. *Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. Suppose an airline conducts a survey. Over Thanksgiving weekend, it surveys six flights from Boston to Salt Lake City to determine the number of babies on the flights. It determines the amount of safety equipment needed by the result of that study.*
a. Using complete sentences, list three things wrong with the way the survey was conducted.
b. Using complete sentences, list three ways that you would improve the survey if it were to be repeated.

Solution a. The survey was conducted using six similar flights.
The survey would not be a true representation of the entire population of air travelers.
Conducting the survey on a holiday weekend will not produce representative results.
b. Conduct the survey during different times of the year.
Conduct the survey using flights to and from various locations.
Conduct the survey on different days of the week.

Exercise 66. *Suppose you want to determine the mean number of students per statistics class in your state. Describe a possible sampling method in three to five complete sentences. Make the description detailed.*

Solution Answers will vary. Sample Answer: Randomly choose 25 colleges in the state. Use all statistics classes from each of the chosen colleges in the sample. This can be done by listing all the colleges together with a two-digit number starting with 00 then 01, etc. The list of colleges can be found on Wikipedia. http://en.wikipedia.org/wiki/List_of_colleges_and_universities_in_California Use a random number generator to pick 25 colleges.

Exercise 67. *Suppose you want to determine the mean number of cans of soda drunk each month by students in their twenties at your school. Describe a possible sampling method in three to five complete sentences. Make the description detailed.*

Solution Answers will vary. Sample Answer: You could use a systematic sampling method. Stop the tenth person as they leave one of the buildings on campus at 9:50 in the morning. Then stop the tenth person as they leave a different building on campus at 1:50 in the afternoon.

Exercise 68. *List some practical difficulties involved in getting accurate results from a telephone survey.*

Solution Answers will vary. Sample Answer: Not all people have a listed phone number. Many people hang up or do not respond to phone surveys.

Introductory Statistics 2nd edition Illowsky Solutions Manual

Exercise 69. *List some practical difficulties involved in getting accurate results from a mailed survey.*

Solution Answers will vary. Sample Answer: Many people will not respond to mail surveys. If they do respond to the surveys, you can't be sure who is responding. In addition, mailing lists can be incomplete.

Exercise 70. *With your classmates, brainstorm some ways you could overcome these problems if you needed to conduct a phone or mail survey.*

Solution Ask everyone to include their age then take a random sample from the data. Include in the report how the survey was conducted and why the results may not be accurate.

Exercise 71. *An instructor takes a sample by gathering data on five randomly selected students from each Lake Tahoe Community College math class. The type of sampling used is*
a. cluster sampling
b. stratified sampling
c. simple random sampling
d. convenience sampling

Solution b

Exercise 72. *A study was done to determine the age, number of times per week, and the duration (amount of time) of residents using a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every eighth house in the neighborhood around the park was interviewed. The sampling method was:*
a. simple random
b. systematic
c. stratified
d. cluster

Solution b

Exercise 73. *Name the sampling method used in each of the following situations:*
a. A person in the airport is handing out questionnaires to travelers asking them to evaluate the airport's service. They do not ask travelers who are hurrying through the airport with their hands full of luggage but instead ask all travelers who are sitting near gates and not taking naps while they wait.
b. A teacher wants to know if students are doing homework, so they randomly select rows two and five and then call on all students in row two and all students in row five to present the solutions to homework problems to the class.

Introductory Statistics 2nd edition Illowsky Solutions Manual

c. The marketing manager for an electronics chain store wants information about the ages of its customers. Over the next two weeks, at each store location, 100 randomly selected customers are given questionnaires to fill out asking for information about age, as well as about other variables of interest.
d. The librarian at a public library wants to determine what proportion of the library users are children. The librarian has a tally sheet on which they mark whether books are checked out by an adult or a child. The librarian records this data for every fourth patron who checks out books.
e. A political party wants to know the reaction of voters to a debate between the candidates. The day after the debate, the party's polling staff calls 1,200 randomly selected phone numbers. If a registered voter answers the phone or is available to come to the phone, that registered voter is asked whom they intend to vote for and whether the debate changed their opinion of the candidates.

Solution

a. convenience
b. cluster
c. stratified
d. systematic
e. simple random

Exercise 74.

A "random survey" was conducted of 3,274 people of the "microprocessor generation" (people born since 1971, the year the microprocessor was invented). It was reported that 48% of those individuals surveyed stated that if they had $2,000 to spend, they would use it for computer equipment. Also, 66% of those surveyed considered themselves relatively savvy computer users.
a. Do you consider the sample size large enough for a study of this type? Why or why not?
b. Based on your "gut feeling," do you believe the percents accurately reflect the U.S. population for those individuals born since 1971? If not, do you think the percents of the population are actually higher or lower than the sample statistics? Why?
Additional information: The survey, reported by Intel Corporation, was filled out by individuals who visited the Los Angeles Convention Center to see the Smithsonian Institute's road show called "America's Smithsonian."
c. With this additional information, do you feel that all demographic and ethnic groups were equally represented at the event? Why or why not?
d. With the additional information, comment on how accurately you think the sample statistics reflect the population parameters.

Solution

a. Yes, in polling, samples that are from 1,200 to 1,500 observations are considered large enough and good enough if the survey is random and is well done.
b. We do not have enough information to decide if this is a random sample from the U.S. population.

Introductory Statistics 2nd edition Illowsky Solutions Manual

c. No, this is a convenience sample taken from individuals who visited an exhibition in the Angeles Convention Center. This sample is not representative of the U.S. population.
d. It is possible that the two sample statistics, 48% and 66% are larger than the true parameters in the population at large. In any event, no conclusion about the population proportions can be inferred from this convenience sample.

Exercise 75. *The Gallup-Healthways Well-Being Index is a survey that follows trends of U.S. residents on a regular basis. There are six areas of health and wellness covered in the survey: Life Evaluation, Emotional Health, Physical Health, Healthy Behavior, Work Environment, and Basic Access. Identify the type of data obtained from each question used in this survey: qualitative, quantitative discrete, or quantitative continuous.*
a. Do you have any health problems that prevent you from doing any of the things people your age can normally do?
b. During the past 30 days, for about how many days did poor health keep you from doing your usual activities?
c. In the last seven days, on how many days did you exercise for 30 minutes or more?
d. Do you have health insurance coverage?

Solution
a. qualitative
b. quantitative discrete
c. quantitative discrete
d. qualitative

Exercise 76. *In advance of the 1936 Presidential Election, a magazine titled Literary Digest released the results of an opinion poll predicting that the republican candidate Alf Landon would win by a large margin. The magazine sent post cards to approximately 10,000,000 prospective voters. These prospective voters were selected from the subscription list of the magazine, from automobile registration lists, from phone lists, and from club membership lists. Approximately 2,300,000 people returned the postcards.*
a. Think about the state of the United States in 1936. Explain why a sample chosen from magazine subscription lists, automobile registration lists, phone books, and club membership lists was not representative of the population of the United States at that time.
b. What effect does the low response rate have on the reliability of the sample?
c. Are these problems examples of sampling error or nonsampling error?
d. During the same year, George Gallup conducted his own poll of 30,000 prospective voters. His researchers used a method they called "quota sampling" to obtain survey answers from specific subsets of the population. Quota sampling is an example of which sampling method described in this module?

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution

a. The country was in the middle of the Great Depression and many people could not afford these "luxury" items and therefore not able to be included in the survey.
b. Samples that are too small can lead to sampling bias.
c. sampling error
d. stratified

Exercise 77.

Crime-related and demographic statistics for 47 US states in 1960 were collected from government agencies, including the FBI's Uniform Crime Report. One analysis of this data found a strong connection between education and crime indicating that higher levels of education in a community correspond to higher crime rates.
Which of the potential problems with samples discussed in Section 1.2 could explain this connection?

Solution

Causality: The fact that two variables are related does not guarantee that one variable is influencing the other. We cannot assume that crime rate impacts education level or that education level impacts crime rate.
Confounding: There are many factors that define a community other than education level and crime rate. Communities with high crime rates and high education levels may have other lurking variables that distinguish them from communities with lower crime rates and lower education levels. Because we cannot isolate these variables of interest, we cannot draw valid conclusions about the connection between education and crime. Possible lurking variables include police expenditures, unemployment levels, region, average age, and size.

Exercise 78.

Imagine you work for a polling company and a member of your team has proposed the following survey question:
"Do you feel happy paying your taxes while some politicians are allowed to use loopholes and avoid paying their fair share of taxes?"
As part of preliminary data collection, 11 people responded to this question. Each participant answered "NO!"
Which of the potential problems with samples discussed in this module could explain this connection?

Solution

Self-Selected Samples: Only people who are interested in the topic are choosing to respond. Sample Size Issues: A sample with only 11 participants will not accurately represent the opinions of a nation.
Undue Influence: The question is wording in a specific way to generate a specific response. Self-Funded or Self-Interest Studies: This question was generated to support one person's claim and it was designed to get the answer that the person desires.

Exercise 79.

A scholarly article about response rates begins with the following quote:

Introductory Statistics 2nd edition Illowsky Solutions Manual

"Declining contact and cooperation rates in random digit dial (RDD) national telephone surveys raise serious concerns about the validity of estimates drawn from such research."
The Pew Research Center for People and the Press admits:
"The percentage of people we interview – out of all we try to interview – has been declining over the past decade or more."
a. What are some reasons for the decline in response rate over the past decade?
b. Explain why researchers are concerned with the impact of the declining response rate on public opinion polls.

Solution

a. Possible reasons: increased use of caller id, decreased use of landlines, increased use of private numbers, voice mail, privacy managers, hectic nature of personal schedules, decreased willingness to be interviewed.
b. When a large number of people refuse to participate, then the sample may not have the same characteristics of the population. Perhaps the majority of people willing to participate are doing so because they feel strongly about the subject of the survey.

Exercise 80.

Fifty part-time students were asked how many courses they were taking this term. The (incomplete) results are shown below:

# of Courses	Frequency	Relative Frequency	Cumulative Relative Frequency
1	30	0.6	
2	15		
3			

Table 1.34 Part-time Student Course Loads

a. Fill in the blanks in Table 1.34.
b. What percent of students take exactly two courses?
c. What percent of students take one or two courses?

Solution

a.

# of Courses	Frequency	Relative Frequency	Cumulative Relative Frequency
1	30	0.6	0.6
2	15	0.3	0.9
3	5	0.1	1.0

b. 30%
c. 90%

Exercise 81.

Sixty adults with gum disease were asked the number of times per week they used to floss before their diagnosis. The (incomplete) results are shown in Table 1.35.

Introductory Statistics 2nd edition Illowsky Solutions Manual

# Flossing per Week	*Frequency*	*Relative Frequency*	*Cumulative Relative Freq.*
0	*27*	*0.4500*	
1	*18*		
3			*0.9333*
6	*3*	*0.0500*	
7	*1*	*0.0167*	

Table 1.35 Flossing Frequency for Adults with Gum Disease

a. Fill in the blanks in Table 1.35.
b. What percent of adults flossed six times per week?
c. What percent flossed at most three times per week?

Solution

a.

# Flossing per Week	**Frequency**	**Relative Frequency**	**Cumulative Relative Freq.**
0	27	0.4500	0.4500
1	18	0.3000	0.7500
3	11	0.1833	0.9333
6	3	0.0500	0.9833
7	1	0.0167	1

Table 1.41

b. 5.00%
c. 93.33%

Exercise 82. *Nineteen immigrants to the U.S were asked how many years, to the nearest year, they have lived in the U.S. The data are as follows: 2; 5; 7; 2; 2; 10; 20; 15; 0; 7; 0; 20; 5; 12; 15; 12; 4; 5; 10.*
Table 1.36 was produced.

Data	*Frequency*	*Relative Frequency*	*Cumulative Relative Frequency*
0	*2*	$\frac{2}{19}$	*0.1053*
2	*3*	$\frac{3}{19}$	*0.2632*
4	*1*	$\frac{1}{19}$	*0.3158*
5	*3*	$\frac{3}{19}$	*0.4737*
7	*2*	$\frac{2}{19}$	*0.5789*
10	*2*	$\frac{2}{19}$	*0.6842*
12	*2*	$\frac{2}{19}$	*0.7895*

Introductory Statistics 2nd edition Illowsky Solutions Manual

15	*1*	$\frac{1}{19}$	*0.8421*
20	*1*	$\frac{1}{19}$	*1.0000*

Table 1.36 Frequency of Immigrant Survey Responses

a. Fix the errors in Table 1.36. Also, explain how someone might have arrived at the incorrect number(s).
b. Explain what is wrong with this statement: "47 percent of the people surveyed have lived in the U.S. for 5 years."
*c. Fix the statement in **b** to make it correct.*
d. What fraction of the people surveyed have lived in the U.S. five or seven years?
e. What fraction of the people surveyed have lived in the U.S. at most 12 years?
f. What fraction of the people surveyed have lived in the U.S. fewer than 12 years?
g. What fraction of the people surveyed have lived in the U.S. from five to 20 years, inclusive?

Solution

a. The Frequencies for 15 and 20 should both be two and the Relative Frequencies should both be $\frac{2}{19}$. The mistake could be due to copying the data down wrong. The Cumulative Relative Frequency for five years should be 0.4737. The mistake is due to calculating the Relative Frequency instead of the Cumulative Relative Frequency. The Cumulative Relative Frequency for 15 years should be 0.8947
b. The 47% is the Cumulative Relative Frequency, not the Relative Frequency.
c. 47% of the people surveyed have lived in the U.S. for five years or less.
d. $\frac{5}{19}$
e. $\frac{15}{19}$
f. $\frac{13}{19}$
g. $\frac{13}{19}$

Exercise 83. *How much time does it take to travel to work? Table 1.37 shows the mean commute time by state for workers at least 16 years old who are not working at home. Find the mean travel time, and round off the answer properly.*

24.0	24.3	25.9	18.9	27.5	17.9	21.8	20.9	16.7	27.3
18.2	24.7	20.0	22.6	23.9	18.0	31.4	22.3	24.0	25.5

Introductory Statistics 2nd edition Illowsky Solutions Manual

24.7	24.6	28.1	24.9	22.6	23.6	23.4	25.7	24.8	25.5
21.2	25.7	23.1	23.0	23.9	26.0	16.3	23.1	21.4	21.5
27.0	27.0	18.6	31.7	23.3	30.1	22.9	23.3	21.7	18.6

Table 1.37

Solution

The sum of the travel times is 1,173.1. Divide the sum by 50 to calculate the mean value: 23.462. Because each state's travel time was measured to the nearest tenth, round this calculation to the nearest hundredth: 23.46.

Exercise 84.

A leading business magazine publishes data on the best small businesses (defined as businesses that have been publicly traded for at least a year, have a stock price of at least $5 per share, and have reported annual revenue between $5 million and $1 billion). Table 1.38 shows the ages of the chief executive officers for the first 60 ranked small businesses.

Age	**Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
40–44	3		
45–49	11		
50–54	13		
55–59	16		
60–64	10		
65–69	6		
70–74	1		

Table 1.38

a. What is the frequency for CEO ages between 54 and 65?
b. What percentage of CEOs are 65 years or older?
c. What is the relative frequency of ages under 50?
d. What is the cumulative relative frequency for CEOs younger than 55?
e. Which graph shows the relative frequency and which shows the cumulative relative frequency?

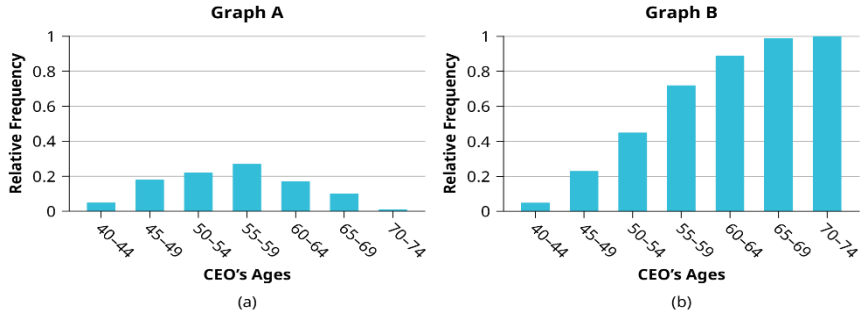


Figure 1.13

Solution

a. 26 (This is the count of CEOs in the 55 to 59 and 60 to 64 categories.)
b. 12% (number of CEOs age 65 or older ÷ total number of CEOs)
c. 14/60; 0.23; 23%

Introductory Statistics 2nd edition Illowsky Solutions Manual

d. 0.45
e. Graph B represents the cumulative relative frequency, and Graph A shows the relative frequency.

Exercise 85. *Table 1.39 contains data on hurricanes that have made direct hits on the U.S. Between 1851 and 2004. A hurricane is given a strength category rating based on the minimum wind speed generated by the storm.*

Category	**Number of Direct Hits**	**Relative Frequency**	**Cumulative Frequency**
1	109	0.3993	0.3993
2	72	0.2637	0.6630
3	71	0.2601	
4	18		0.9890
5	3	0.0110	1.0000
	Total = 273		

Table 1.39 Frequency of Hurricane Direct Hits

What is the relative frequency of direct hits that were category 4 hurricanes?
a. 0.0768
b. 0.0659
c. 0.2601
d. Not enough information to calculate

Solution b

Exercise 86. *Table 1.39 contains data on hurricanes that have made direct hits on the U.S. Between 1851 and 2004. A hurricane is given a strength category rating based on the minimum wind speed generated by the storm.*

Category	**Number of Direct Hits**	**Relative Frequency**	**Cumulative Frequency**
1	109	0.3993	0.3993
2	72	0.2637	0.6630
3	71	0.2601	
4	18		0.9890
5	3	0.0110	1.0000
	Total = 273		

Table 1.39 Frequency of Hurricane Direct Hits

What is the relative frequency of direct hits that were AT MOST a category 3 storm?
a. 0.3480
b. 0.9231
c. 0.2601
d. 0.3370

Introductory Statistics 2nd edition Illowsky Solutions Manual

Solution b

Exercise 87. *How does sleep deprivation affect your ability to drive? A recent study measured the effects on 19 professional drivers. Each driver participated in two experimental sessions: one after normal sleep and one after 27 hours of total sleep deprivation. The treatments were assigned in random order. In each session, performance was measured on a variety of tasks including a driving simulation.*
Use key terms from this module to describe the design of this experiment.

Solution Explanatory variable: amount of sleep
Response variable: performance measured in assigned tasks
Treatments: normal sleep and 27 hours of total sleep deprivation
Experimental Units: 19 professional drivers
Lurking variables: none – all drivers participated in both treatments
Random assignment: treatments were assigned in random order; this eliminated the effect of any "learning" that may take place during the first experimental session
Control/Placebo: completing the experimental session under normal sleep conditions
Blinding: researchers evaluating subjects' performance must not know which treatment is being applied at the time

Exercise 88. *An advertisement for Acme Investments displays the two graphs in Figure 1.14 to show the value of Acme's product in comparison with the Other Guy's product. Describe the potentially misleading visual effect of these comparison graphs. How can this be corrected?*

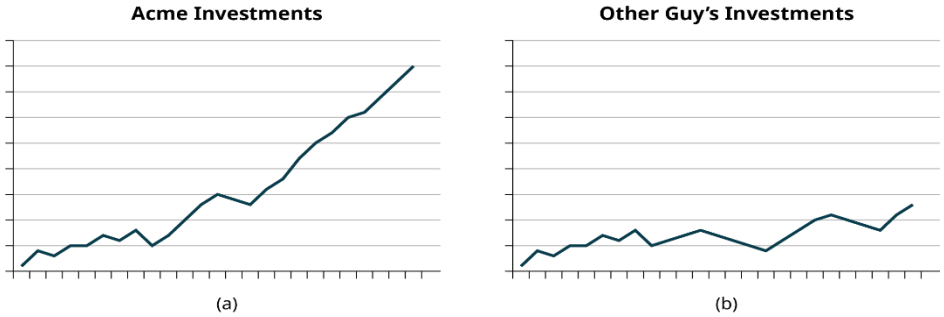


Figure 1.14 As the graphs show, Acme consistently outperforms the Other Guys!

Solution The graphs do not show scales of values. We do not know the period of time each graph represents; they may show data from different years. We also do not know if the vertical scales on each graph are equivalent. The scales may have been adjusted to exaggerate or minimize trends. There is no reliable information to be gleaned from these graphs, and setting them up as examples of performance is misleading.

Introductory Statistics 2nd edition Illowsky Solutions Manual

Exercise 89. *The graph in Figure 1.15 shows the number of complaints for six different airlines as reported to the US Department of Transportation in February 2013. Alaska, Pinnacle, and Airtrain Airlines have far fewer complaints reported than American, Delta, and United. Can we conclude that American, Delta, and United are the worst airline carriers since they have the most complaints?*

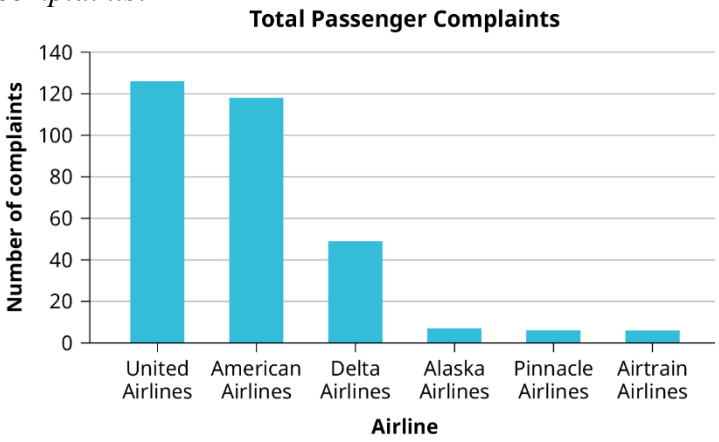


Figure 1.15

Solution You cannot assume that the numbers of complaints reflect the quality of the airlines. The airlines shown with the greatest number of complaints are the ones with the most passengers. You must consider the appropriateness of methods for presenting data; in this case displaying totals is misleading.

Exercise 90. *Seven hundred and seventy-one distance learning students at Long Beach City College responded to surveys in a specific academic year. Highlights of the summary report are listed in Table 1.40.*

Have computer at home	96%
Unable to come to campus for classes	65%
Age 41 or over	24%
Would like LBCC to offer more DL courses	95%
Took DL classes due to a disability	17%
Live at least 16 miles from campus	13%
Took DL courses to fulfill transfer requirements	71%

Table 1.40 LBCC Distance Learning Survey Results

a. What percent of the students surveyed do not have a computer at home?
b. About how many students in the survey live at least 16 miles from campus?
c. If the same survey were done at Great Basin College in Elko, Nevada, do you think the percentages would be the same? Why?

Solution a. 4%
b. 13%

Introductory Statistics 2nd edition Illowsky Solutions Manual

c. Not necessarily. Long beach City is the seventh largest in California the college has an enrollment of approximately 27,000 students. On the other hand, Great Basin College has its campuses in rural northeastern Nevada, and its enrollment of about 3,500 students.

Exercise 91. *Several online textbook retailers advertise that they have lower prices than on-campus bookstores. However, an important factor is whether the Internet retailers actually have the textbooks that students need in stock. Students need to be able to get textbooks promptly at the beginning of the college term. If the book is not available, then a student would not be able to get the textbook at all, or might get a delayed delivery if the book is back ordered.*
Martin, a college newspaper reporter, is investigating textbook availability at online retailers. He decides to investigate one textbook for each of the following seven subjects: calculus, biology, chemistry, physics, statistics, geology, and general engineering. He consults textbook industry sales data and selects the most popular nationally used textbook in each of these subjects. He visits websites for a random sample of major online textbook sellers and looks up each of these seven textbooks to see if they are available in stock for quick delivery through these retailers. Based on his investigation, Martin writes an article in which he draws conclusions about the overall availability of all college textbooks through online textbook retailers. Write an analysis of his study that addresses the following issues: Is his sample representative of the population of all college textbooks? Explain why or why not. Describe some possible sources of bias in this study, and how it might affect the results of the study. Give some suggestions about what could be done to improve the study.

Solution Answers will vary. Sample answer: The sample is not representative of the population of all college textbooks. Two reasons why it is not representative are that he only sampled seven subjects and he only investigated one textbook in each subject. There are several possible sources of bias in the study. The seven subjects that he investigated are all in mathematics and the sciences; there are many subjects in the humanities, social sciences, and other subject areas, (for example: literature, art, history, psychology, sociology, business) that he did not investigate at all. It may be that different subject areas exhibit different patterns of textbook availability, but his sample would not detect such results.
He also looked only at the most popular textbook in each of the subjects he investigated. The availability of the most popular textbooks may differ from the availability of other textbooks in one of two ways:
• the most popular textbooks may be more readily available online, because more new copies are printed, and more students nationwide are selling back their used copies OR
• the most popular textbooks may be harder to find available online, because more student demand exhausts the supply more quickly.

Introductory Statistics 2nd edition Illowsky Solutions Manual

In reality, many college students do not use the most popular textbook in their subject, and this study gives no useful information about the situation for those less popular textbooks.
He could improve this study by:
• expanding the selection of subjects he investigates so that it is more representative of all subjects studied by college students, and
• expanding the selection of textbooks he investigates within each subject to include a mixed representation of both the most popular and less popular textbooks.

This file is copyright 2023, Rice University. All Rights Reserved.

Introductory Statistics 2nd edition Illowsky Solutions Manual