Solutions Manual for Business Statistics Communicating with Numbers, 4e Sanjiv Jaggia, Alison Kelly

written by

tutorsection

www.stuvia.com

Chapter 1. Data and Data Preparation

Solutions

1.
   a. The population is all iPhone 4 users.

   b. Sample statistics

2. The value 35 is the estimated average age of the population. It is both costly and time consuming (likely impossible) to take a census of all video game players and compute the actual average age.

3.
   a. The population is all students enrolled in the accounting class.

   b. The value 3.29 represents the population parameter since we are not choosing a sample but drawing results from the actual population.

4.
   a. The population is all recent college graduates with an engineering degree.

   b. No, the average salary is a sample statistic computed from a sample, not the population.

5.
   a. The population is all elderly people. The sample consists of 949 elderly people.
   b. 22% and 17% represent sample statistics.

6. The data are cross-sectional data.
Note: Data will vary due to the nature of sampling.

7. The data are time series data.
Note: Data will vary depending on the date of retrieval. These data were retrieved July 2019.

Date	Adj. Close Price
7/1/2018	192.55
8/1/2018	195.72
9/1/2018	202.97
10/1/2018	172.33
11/1/2018	176.68
12/1/2018	169.36
1/1/2019	180.90
2/1/2019	182.49
3/1/2019	189.15
4/1/2019	202.28
5/1/2019	188.53
6/1/2019	206.52

8. The front page of the New York Times website is likely to be textual (written reports) with multimedia contents (photographs, etc.). The resulting data are unstructured in that they do not conform to a predefined row-column format.

9. Data on price and fuel economy of small hybrid vehicles can be specified in a predefined row-column format, and therefore, are structured.
Note: Data will vary due to the nature of sampling.

10. The data for Under Armour's annual revenue are structured since they are specified in a well-defined row-column format. The data are time series data.
Note: Data will vary depending on the date of retrieval. These data were retrieved July 2019.

Year	Annual Revenue (in $ millions)
2018	$5,193
2017	$4,989
2016	$4,833
2015	$3,963
2014	$3,084
2013	$2,332

2012	$1,835
2011	$1,473
2010	$1,064
2009	$856

11. The resulting data about online social media usage have a well-defined row-column format, and therefore, are structured. The data are cross-sectional data.
Note: Data will vary due to the nature of sampling.

12.
a. Numerical; discrete
b. Categorical
c. Numerical; continuous

13.
a. Categorical
b. Numerical; continuous
c. Numerical; discrete

14.
a. Nominal
b. Interval
c. Ordinal

15.
a. Ratio
b. Ordinal
c. Nominal

16.
a. Ratio
b. Interval
c. Ratio

17.
a. Nominal scale; the values differ in name.
b.

Major	# of Students

Accounting	4
Economics	4
Finance	4
History	4
Management	5
Psychology	3
Statistics	3
Undecided	3

c. An inspection of the data shows that Management has the highest number of students whereas Psychology, Statistics, and Undecided have the lowest.

18.

a. Nominal
b. Interval. The observations for Year can be ranked, categorized and measured when using this kind of scale. However, there is no true zero point so we cannot calculate meaningful ratios between years.
c. Ratio. This type of scale is the strongest form of measurement. There is a true zero point which allows for the calculation of meaningful ratios between observations.

19.

a. 50 observations of x_2 are equal to two.
b. When sorted in ascending order the first observation is $x_1 = 0$ and $x_2 = 1$.
c. When sorted in descending order the first observation is $x_1 = 1$ and $x_2 = 2$.
d. When x_1 is sorted in ascending order and x_2 is sorted in descending order, the first observation is $x_1 = 0$ and $x_2 = 1$.
e. 0 values are missing for x_1 and x_2.

20.

a. 17 observations of x_1 are greater than 30.
b. When sorted in ascending order the first observation is $x_1 = 2$, $x_2 = 540$, and $x_3 = 201$.
c. When x_1 and x_2 are sorted in descending order, and x_3 is sorted in ascending order, the first observation is $x_1 = 52$, $x_2 = 218$, and $x_3 = 154$.
d. 1 value is missing for x_1, 3 values are missing for x_2, and 5 values are missing for x_3.

21.

a. 85 observations of x_4 are less than 3.

b. When sorted in ascending order, the first observation is $x_1 = 0.0121$, $x_2 = 24$, $x_3 = 27$, and $x_4 = 3$.
c. When sorted in descending order, the first observation is $x_1 = 0.9980$, $x_2 = 196$, $x_3 = 13$, and $x_4 = 3$.
d. 0 values are missing for x_1, x_2, x_3, and x_4.
e. The categorical variable x_4 has three categories: 1, 2 and 3. Category 1 has 40 observations, category 2 has 45 observations, and category 3 has 33 observations.

22.

a. There are 25 observations in the subset.

x_1	x_2	x_3	x_4
544	616	5/14/85	1
⋮	⋮	⋮	⋮
186	218	3/8/81	1

b.

x_1	x_2	x_3	x_4
544	616	5/14/85	1
351	433	2/10/81	1
…	…	…	…
466	387	7/6/84	1

There are 24 observations for $x_4 = 1$.

x_1	x_2	x_3	x_4
408	348	10/27/99	0
533	576	3/3/93	0
…	…	…	…
202	170	6/22/94	0

There are 28 observations for $x_4 = 0$.

23.

a.

x_2	x_3	x_4
Own	189	33
Rent	120	28
…	…	…
Own	136	23

There are no missing values

Thank you for your order,
Copy the link below in your internet browser to download your order file complete.

https://www.mediafire.com/file/e9fnf6n16mp6l7w/SM_Business_Statistics_Communicating_with_Numbers%252C_4e_Sanjiv_Jaggia%252C_Alison_Kelly.zip/file

I would appreciate it if you could leave positive feedback. That would be helpful for my store, and I will surely leave positive feedback as well!! Thank you, and have a wonderful day :))

b.

x_2	x_3	x_4
Rent	120	28
Rent	128	23
…	…	…
Rent	128	13

35 observations remain.

24.

a. Variables x_1, x_2 and x_4 have at least one missing value.
b. Observations 13, 18, 21, 25, 35, 40, and 48 have missing values.
c. 45 observations remain after missing values are removed.

25.

a. Minnesota has the highest average writing score of 643. The average math score of Minnesota is 655.
b. The Virgin Islands has the lowest average math score of 445. The average writing score of the Virgin Islands is 490.
c. 13 states reported average math scores above 600.
d. 25 states reported average writing scores below 550.

26.

a. 1 of the 10 highest income earners is married and always exercises.
b. 9 individuals are married, exercise sometimes, and earn more than $110,000 per year.
c. 5 values are missing for Exercise, 2 for Marriage, and 3 for Income.
d. 281 individuals are married, and 134 are not.
e. 69 married individuals always exercise, and 74 unmarried individuals never exercise.

27.

a. In the data there are 508 males and 382 females.
b. $\frac{336}{508} \times 100\% = 66.1417\%$ of males in the data are married, and $\frac{248}{382} \times 100\% = 64.9215\%$ of the females in the data are married.
c. Of the 10 individuals with the highest income 7 are married males.
d. The largest income among females is $138,000, the smallest is $22,000. The largest income among males is $147,000, the smallest is $35,000.

e. The largest income among married males is $147,000, the smallest is $35,000. The largest income among unmarried males is $140,000, the smallest is $39,000.

28.

a. There are missing values in the variable Travel Plan. All other variables are complete.
b. 300 observations are removed.
c.

College	FoodSpend	Income	TravelPlan
No	1,892.37	77,626	1
No	6,865.77	77,626	1

2 observations remain in the subset.

29.

a. 2018 Population ≥ 5,000,000: 23 observations in this subset.
2018 Population < 5,000,000: 27 observations in this subset.
b. 9 states are removed.

30.

a. Yes, there are missing values in the data set, the variable Yards is missing for observation 25, Attempts for observation 28, and Interceptions for observation 29. All other variables and observations are complete.
b. 3 observations were removed using the omission strategy.
c. 4 of the observations had less than 5 touchdowns and were removed, additionally 1 of the observations had more than 20 interceptions and were removed. A total of 5 observations were removed from this process.

31.

a. 4 subsets were created: “Dept of HR”, “IT Dept”, “Public Utilities”, and “Sustainability and Environ Dept”.
b. There are 2 missing values in the “IT Dept” Hourly rate variable, all other subsets and variables are complete.

32.

a. The relevant population consists of all US citizens born in 2019.
b. The averages are computed from the relevant samples.

Thank you for your order,
Copy the link below in your internet browser to download your order file complete.

https://www.mediafire.com/file/e9fnf6n16mp6l7w/SM_Business_Statistics_Communicating_with_Numbers%252C_4e_Sanjiv_Jaggia%252C_Alison_Kelly.zip/file

I would appreciate it if you could leave positive feedback. That would be helpful for my store, and I will surely leave positive feedback as well!! Thank you, and have a wonderful day :))

33. The resulting data about living accommodations and expenses have a well-defined row-column format, and therefore, are structured. The data are cross-sectional data.
Note: Data will vary due to the nature of sampling.

34.

State	Median Household Income in 2013-2017
Alabama	46,472
Arizona	53,510
California	67,169
Florida	50,883
Georgia	52,977
Indiana	52,182
Iowa	56,570
Maine	53,024
Massachusetts	74,167
Minnesota	65,699
Mississippi	42,009
New Mexico	46,718
North Dakota	61,285
Washington	66,174

Source: *http://www.census.gov/; Retrieved July 1, 2019.*

These data are estimates for the period 2013-2017. The data are cross-sectional data. In this particular group of states, Massachusetts has the highest median income, whereas Mississippi has the lowest median income. Also, states in the North such as Massachusetts, Minnesota and Washington tend to have higher incomes than Southern states.

35.

a. The variable Sex is categorical where as Income, Age, and Spending are numerical.
b. The measurement scale is nominal for Sex and ratio for Income, Age, and Spending. Recall, that the nominal and ratio scales represent the least and most sophisticated levels of measurement, respectively.

36.

a. The variable Year is measured on the interval scale because the observations can be ranked, categorized and measured when using this kind of scale. However, there is no true zero point so we cannot calculate meaningful ratios between years.

b. The variable Quarter is measured on the nominal scale, even though it contains numbers. It is the least sophisticated level of measurement because if we are presented with nominal data, all we can do is categorize or group the data.
c. The variable Vacation is measured on the ratio scale. It is the strongest level of measurement because it allows us to categorize and rank the data as well as find meaningful differences between observations. Also, with a true zero point, we can interpret the ratios between observations.

37.

a. The first customer in this list spent $3,362.86 on food.
b. Of the 10 customers who spent the most of travel, 3 were home owners, and 1 was both a home and car owner.
c. 3 values are missing for the variable OwnHome, 3 values are missing for the variable OwnCar, 0 values are missing for the variable FoodSpend, and 1 value is missing for the variable TravelSpend.
d. 133 of the customers own homes. 88 of the customers own homes but do not own cars.

38.

a. In the data there are 616 males and 614 females.
b. $\frac{304}{616} \times 100\% = 49.3506\%$ of males in the data, and $\frac{352}{614} \times 100\% = 57.3290\%$ of the females in the data are admitted.
c. 2 of the 10 students with the highest high school GPAs are male.
d. 6 of the 10 students with the lowest SAT score are female.
e. The highest SAT score of admitted males is 1600, the lowest is 1055. The highest SAT score of admitted females is 1599, the lowest is 1062.

39.

a. Observation 26 is missing for the variable Siblings, observation 13 is missing for the variable Height, observation 47 is missing for the variable Weight, and observations 17 and 51 are missing for the variable Income. All other variables are complete.
b. A total of 5 missing values are removed.

40.

a. All variables other than Name have missing values. Observation 20 is missing for the variable Price; observation 173 is missing for variable Dividend; observations 38 is missing for the variable PE; observation 26 is missing for variable EPS, observation 98 is missing for the variable Book

Value; observation 26, and 154 are missing for the variable 52 week low, observation 46 is missing for the variable 52 week high, and observation 51 is missing for variable the EBITDA. There are 9 missing values in the data set.

b. There are 367 complete observations in the subset.
c. 61 observations remain in the data set for which EPS is less than 15.