

Chapter 1: Introduction to Data

SECTION 1.2

- 1.1 a. The variable “Handedness” is a categorical variable since there are categories of “Left” or “Right.”
 b. The variable “Age” is a numerical variable since it is quantitative.
- 1.2 a. The variable “Shoe size” is a numerical variable since it is quantitative.
 b. The variable “Eye color” is a categorical variable since it has categories such as “Blue” and “Brown.”
- 1.3 Persons who have responded that they are male have been coded with a “1,” and the others are coded with a “0”. Since these are categories, the variable “Male” is categorical. Yes, summing the “Male” categorical variable makes sense; doing so gives the total number of persons who classified themselves as males.
- 1.4 We use a “1” for full time students, and for those who are not are classified as full time, we use “0.”

Units	Full
16.0	1
13.0	1
5.0	0
15.0	1
19.5	1
11.5	0
9.5	0
8.0	0
13.5	1
12.0	1
14.0	1

- 1.5 a. The data are presented in a stacked format.
 b. Males are indicated by “1” and females are indicated t by “0.”
 c. The table shows the data presented in an unstacked format.

Female	Male
9.5	9.4
9.5	9.5
9.9	9.5
	9.7

- 1.6 a. The data are presented in an unstacked format.
 b. The table shows the data presented in a stacked format.

Age	Class
31	5 PM
34	5 PM
46	5 PM
47	5 PM
50	5 PM
24	Noon
18	Noon
21	Noon
20	Noon
20	Noon

SECTION 1.3

- 1.7** **a.** The percentage of answers that were changed for the students told not to change is found by dividing the number of students who changed by the total number of students who were told not to change
This gives $\frac{189}{29617} \approx 0.00638$ or approximately 0.64%.
- b.** The percentage of answers that were changed for the students told to change is found by dividing the number of students who changed by the total number of students who were told not to change. This gives $\frac{124}{14513} \approx 0.00854$ or approximately 0.85%.
- c.** The higher percentage of students who changed is found in the class that was told to change their answers if they felt they had found a better answer, so it appears that perhaps the instruction to do so had a measurable effect.
- 1.8** **a.** For the class that was instructed not to change, a total of 189 students changed their answers. The percentage of changes made from wrong to right is therefore $\frac{91}{189} \approx 0.481$ or approximately 48.1%. The percentage of changes made from right to wrong is $\frac{48}{189} \approx 0.254$ or 25.4%. Since more students changed from wrong to right, then on average their grades improved.
- b.** For the class that was instructed to change, a total of 124 students changed their answers. The percentage of changes made from wrong to right is then $\frac{86}{124} \approx 0.694$ or 69.4%. The percentage of changes made from right to wrong is $\frac{24}{124} \approx 0.194$ or 19.4%. Since more students changed from wrong to right, then on average their grades improved.
- c.** It is better to change your answer if you believe the second one you found is more reasonable.
- 1.9** **a.** This statistics class has a total of 38 students, so the percentage that is male is $\frac{15}{38} \approx 0.395$ or 39.5%.
- b.** 64.1% of 234 can be calculated by $0.641 \times 234 = 149.994$. Therefore, it is likely that 150 members of the class are men.
- c.** This time the class size is unknown, call it C. Then 40% of C is 20 or in symbols:
 $0.40 \times C = 20$ or equivalently $C = \frac{20}{0.4} = 50$.
- 1.10** **a.** The number of nurses who are male is 35% of 346. Calculate this by $0.35 \times 346 = 121.1$. So, 121 of the nurses are male.
- b.** The number of female engineers is $178 - 112 = 66$, so the percentage of engineers who are female is $\frac{66}{178} \approx 0.371$ or 37.1%.
- c.** The total number of lawyers in the firm is unknown; let's call it T. Then we know that 65% of T is 169. In symbols this is:
 $0.65 \times T = 169$ or equivalently $T = \frac{169}{0.65} = 260$
- 1.11** The frequency of women is seven. The proportion of women is $\frac{7}{11}$ and the percentage is approximately 63.6%.
- 1.12** The frequency of right-handed people is nine. The proportion who are right-handed is $\frac{9}{11}$ or approximately 81.8%.

- 1.13** If we let T represent the total predicted U.S. population in 2050, then we know that 20.2% of T is 88,547,000. In symbols we have

$$0.202 \times T = 88,547,000 \text{ or equivalently } T = 88,547,000 / 0.202 \approx 438,351,485$$

Thus, there will be 438,351,485 persons living in the U.S. in 2050. You may round this answer.

- 1.14** If we let T represent the total number of persons living in the U.S. in 2007, then we are told that there were 12,608,000 of these persons, representing 5.5% of the total U.S. population. Thus, 5.5% of T is 12,608,000 or in symbols

$$0.055 \times T = 12,608,000 \text{ or equivalently } T = 12,608,000 / 0.055 \approx 229,236,364$$

Rounding up gives 229,236,364 U.S. residents. You may round this answer.

- 1.15** Calculate the percentage of persons living with AIDS by dividing the number of persons with AIDS in each state by that state's total population. For example, to find the percentage of persons with AIDS in New York, compute $75,253 / 19,297,729 \approx 0.0039$. Then the rate of AIDS infection per thousand is $0.0039 \times 1000 = 3.90$. The other calculations are made similarly. The table shows the resulting data values.

State	AIDS	Rank by Cases	Population	Percentage With AIDS	AIDS per 1000	Rank by Rate
New York	75,253	1	19,297,729	0.003900	3.90	2
California	65,582	2	36,553,215	0.001794	1.79	4
Florida	48,059	3	18,251,243	0.002633	2.63	3
Texas	34,940	4	23,904,380	0.001462	1.46	6
Pennsylvania	19,236	5	12,432,792	0.001547	1.55	5
District of Columbia	8,895	6	588,292	0.015120	15.12	1

The rankings based on the rate of AIDS infection are different than the rankings based on the number of AIDS cases. For example, District of Columbia has the lowest number of AIDS cases but has the highest rate of infection.

Among the five states and the District of Columbia, you would be most likely to meet someone with AIDS if you were in the District of Columbia and least likely if you were in Texas.

- 1.16** Calculate the population density by dividing the population in each state by the total area of that state. For example, to find the population density of Pennsylvania, compute $12,448,279 / 44,817 \approx 277.76$. The other calculations are made similarly. The table shows the resulting data values.

State	Population	Area	Population Density	Rank by Density
Pennsylvania	12,448,279	44,817	277.76	3
Illinois	12,901,563	55,584	232.11	5
Florida	18,328,340	53,927	339.87	2
New York	19,490,297	47,214	412.81	1
Texas	24,326,974	261,797	92.92	6
California	36,756,666	155,959	235.68	4

New York is the most densely populated state and Texas is the least densely populated state.

- 1.17** Simply compute the percentage as the number of married adults divided by the total number of adults in the U.S. The table shows the resulting data values.

Married	Total	Percentage
112.6	191.8	58.7
116.8	207.2	56.4
120.2	213.8	56.2
129.9	235.8	55.1

The percentage of married persons is decreasing over time.

- 1.18** Simply compute the percentage of elderly as the number of elderly adults divided by the total number of adults in the U.S. The table shows the resulting data values.

Year	Elderly	Population	Percentage
1900	3,080	75,100	4.10
1920	4,933	105,000	4.70
1940	9,019	133,000	6.78
1960	16,560	180,000	9.20
1980	25,549	226,000	11.30
2000	34,992	282,000	12.41

Elderly adults represent an ever-increasing portion of the population.

- 1.19** We do not know the total number of students in each class. The smaller number of females in the later class may simply be because there are fewer total students in that class.
- 1.20** No, you cannot conclude that pedestrians are safer in Hillsborough County, because we do not know how many persons live in this county versus Miami-Dade. It is possible that there are more pedestrians in Miami-Dade County resulting in more fatalities.

SECTION 1.4

- 1.21** This is an observational study since subjects have not been randomly assigned to either of a treatment or comparison group.
- 1.22** This is an observational study since subjects have not been randomly assigned to either of a treatment or comparison group.
- 1.23** This is a controlled experience since subjects have been randomly assigned to either the treatment (drinking one before the other) or comparison group (drinking in the other order).
- 1.24** This is a controlled experience since subjects have been randomly assigned to either the treatment (new drug) or comparison group (placebo).
- 1.25** This is a controlled experience since subjects have been randomly assigned to either the treatment (cartoons with violence) or comparison group (cartoons without violence).
- 1.26** This is an observational study since subjects have not been randomly assigned to either of a treatment or comparison group. Student responses have simply been observed.

- 1.27** This is an observational study since subjects have not been randomly assigned to either of a treatment or comparison group. Student actions have simply been observed.
- 1.28** This is a controlled experience since subjects have been randomly assigned to either the treatment (exercise running stairs) or comparison group (no exercise).
- 1.29** No, this does not necessarily imply that the tutoring resulted in higher grades. This is not a controlled experiment since students were allowed to choose tutoring or not. A possible confounding variable would be “motivation.” The same motivation that moves students to seek extra assistance in the form of tutoring may also result in higher grades.
- 1.30**
- a.** This action destroys the randomness that is critical to the design of any statistical study and introduces bias into the experiment.
 - b.** Study participants must be randomly assigned to either treatment or control group.
 - c.** This will likely introduce bias, for the doctor may well give higher evaluations to those receiving antidepressants.
 - d.** The experiment should be made a double-blind study in which neither participants nor researchers know to which group (treatment or control) participants are assigned.
- 1.31** This was an observational study since participants were not randomly assigned to either receiving stents or CABG. Researchers cannot conclude that CABG results in lower fatalities or fewer serious complications because it is possible that there are confounding variables, other factors shared by those receiving CABG that account for their improvement.
- 1.32** This is most likely an observational study since it appears individuals were simply observed as to their activity level and level of cognition. Researchers cannot conclude that exercise reduces risk to cognition loss because there may well be confounding variables.
- 1.33** The respondents should be randomly assigned to one or the other of the two (treatment or control) groups. . Each group’s participants should receive a pill: vitamin C for those in the treatment group and a placebo for those in the control group. Furthermore, neither students (nor researchers) should know which group the students are placed in. This would make it a double-blind study.
- 1.34**
- a.** Those who chose to be a part of the vegetarian (treatment) group may be more likely to make healthy choices around exercise and personal wellness. It may be these other factors that lead to weight loss and not the diet itself.
 - b.** The study should be redesigned so that it ensures participants are *randomly* assigned to one or the other of the treatment or control groups.
- 1.35** No, this is not a randomized experiment. As such, we cannot conclude that optimism lowers the risk of developing the common cold. Other confounding variables may account for the lower risk. For example, in general, healthier individuals who exercise will likely have stronger immunity; these are the same kinds of individuals who would be expected to score high on “energy, happiness, and relaxation.”
- 1.36** This study does not include a control group, whose participants would receive a placebo instead of melatonin.
- 1.37** The only aspect that is unclear is whether participants were assigned at random to one or the other of the two groups. Because randomization limits bias, without it we could not conclude that smaller doses are as effective as standard doses.
- 1.38** The only aspect that is unclear is whether participants were assigned at random to one or the other of the three groups. Because randomization limits bias, without it we could not conclude that the presence of confederates effects compliance.

- 1.39** This is an observational study since participants were not assigned at random to either a treatment (extra vitamin C) group or control (lower vitamin C level) group. We cannot conclude vitamin C lowers allergy risk to infants because there may be other factors associated with taking vitamin C to which lower allergy risk may be attributed.
- 1.40** This is most likely an observational study. If it were a controlled experiment, researchers would need child-age participants whose parents would willingly allow them to be subjected to second-hand smoke; this is highly unlikely.
- 1.41**
- a.** Fifty mice are in each group. The percentage with tumors from the LD group is $\frac{4}{50} = 0.08$ or 8% while the percentage with tumors from the LL group is $\frac{14}{50} = 0.28$ or 28%. It appears that continuous exposure to light may cause the formation of tumors.
 - b.** This was a controlled experiment since mice were randomly assigned to one of the two groups.
 - c.** Yes, researchers can conclude that continuous exposure to light causes tumors because the random assignment of mice ensures that the effect of confounding variables and biases are minimized.
- 1.42**
- a.** Of the 57 children total who attended preschool, 37 later graduated high school. So, $\frac{37}{57}$ or 64.9% of children who attended preschool later graduated high school.
 - b.** Of the 64 children total who did not attend preschool, 29 later graduated high school. So, $\frac{29}{64}$ or 45.3% of children who did not attend preschool later graduated high school.
 - c.** Yes, early educational experiences have proven to have a variety of benefits.
 - d.** Yes, since this was a controlled experiment in which participants were randomly assigned to one or the other of the two (treatment or control) groups, bias is limited and the researchers may conclude that preschool attendance positively effects high school graduation rates.

CHAPTER REVIEW EXERCISES

- 1.43**
- a.** 81 of the 440 persons dating or 18.4% were obese.
 - b.** 103 of the 429 persons cohabitating or 24% were obese.
 - c.** 147 of the 424 married persons or 34.7% were obese.
 - d.** The group of married persons had the highest rate of obesity. Because this is not a controlled experiment, we cannot conclude that a person's marital status directly affects that person's likelihood of being obese. Age may be a confounding variable since older persons are more likely to be married (than cohabitating or dating) and also overweight. Other confounding variables are possible.
- 1.44**
- a.** This is an observational study. If it were a controlled experiment, researchers would have assigned participants at random to one of either the treatment or the control groups.
 - b.** Because it is only an observational study, we cannot infer that prolonged head set use causes brain cancer. Nor can we conclude from this study that cell phones are more dangerous than cigarettes. From the statistics given in the study, the fatality rate due to smoking cigarettes is approximately 5 per 1000 (5 million die per 1 billion who smoke). However, given only the fact that prolonged head set use doubles the risk of developing brain cancer, we have no way to determine a fatality rate from cell phone use.

- 1.45** **a.**

	Girl	Boy	Totals
Violent	11	10	21
Nonviolent	4	19	23
Totals	15	29	44

- b. The percentage of girls on probation for violent crimes is $\frac{11}{15}$ or 73.3%.
The percentage of boys on probation for violent crimes is $\frac{10}{29}$ or 34.5%.
- c. Girls are more likely to have participated in violent crimes. In fact, the percentage of girls on probation for violent crimes is more than double the percentage of boys on probation for similar crimes.

1.46

	Antivenom	Placebo	Totals
Better	7	1	8
Not Better	1	6	7
Totals	8	7	15

The percentage of those receiving antivenom who are later classified as better is $\frac{7}{8}$ or 87.5%.

The percentage of those receiving the placebo who are later classified as better is $\frac{1}{7}$ or 14.3%.

The percentage of those improving while on antivenom is over six times as high as the percentage of those improving while on placebo.

1.47 Answers will vary. A sample answer is given below.

As a researcher, you should create two groups, a treatment group of persons to receive activated charcoal and a control group of persons to receive a placebo. Assign the 200 participants who are suffering from “toxin disease” at random to one or the other of the two groups. Ideally, neither the subjects nor the researchers should know to which group they have been assigned (a so-called double-blind study).

1.48 Answers will vary. A sample answer is given below.

As a researcher, you should create two groups, a treatment group of persons to receive grapeseed extract and a control group of persons to receive a placebo. Assign the 200 participants who are suffering from allergies at random to one or the other of the two groups. Ideally, neither the subjects nor the researchers should know to which group they have been assigned (a so-called double-blind study).

1.49 This is not a controlled experiment but rather an observational study. As a result, we cannot conclude that the lower fatality rate is due to willingness to participate in a clinical trial. Other factors may well account for the significantly lower fatality rate. For example, perhaps those who are willing to participate in a clinical trial are also more motivated to engage in healthy behaviors that may prolong life.

1.50 This is not a controlled experiment but rather an observational study. As a result, we cannot conclude that the lower recidivism rate is due to the comfortable prison accommodations. Other factors may well account for the much lower recidivism rate for those previously incarcerated in Halden Fengsel prison.

1.51 This is not a controlled experiment but rather an observational study. There is no control group containing participants who have not had significant brain injuries for comparison. The rate of violent crime resulting in death row sentences from the treatment group and a control group must be evaluated for comparison. Without such a comparison, we cannot conclude that head trauma causes higher rates of violent crime eventually resulting in death row sentences.

1.52 This is not a controlled experiment but rather an observational study. There is no control group containing participants who did not engage in brief high-intensity exercise for comparison. The results of diabetes tests from the treatment group and a control group must be compared. Without such a comparison, we cannot conclude that brief high-intensity lowers diabetes indicators.

