Elementary Statistics

FIFTH EDITION

NEIL A. WEISS

Student's Solutions Manual

DAVID LUND

STUDENT'S SOLUTIONS MANUAL

David Lund
University of Wisconsin - Eau Claire

Elementary Statistics

Fifth Edition

NEIL A. WEISS

Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Reproduced by Addison-Wesley from camera-ready copy supplied by the author.

Printed in the United States of America.

ISBN 0-201-73471-0

3 4 5 6 7 8 9 10 VG 04 03

PREFACE

Student's Solutions Manual is designed to be used with the text *Elementary Statistics, Fifth Edition* by Neil A. Weiss. It provides complete solutions to the 532 odd-numbered exercises and all of the 300 review problems in Chapters 1-14 of the text. The solutions are more than answers--intermediary steps in the process of solving the exercises are also provided.

New to this edition of the text are 510 supplementary exercises which are on the Weiss Stats CD supplied with the text. In this manual, we have also provided the answers to all 259 of the odd-numbered supplementary exercises and complete solutions for some of them.

Note: Many of the numerical answers presented here were obtained by using computer software and the original set of data. If you solve problems by hand and do some intermediate rounding or use the summary statistics which are provided in the text for many of the exercises, your answers may differ slightly from the ones given in this manual. In a few instances, the solutions provided here may also differ slightly from the answers provided in Appendix B of the text, but we are not aware of any instances in which the differences are significant enough to alter the conclusions reached. As noted by the author, you should retain intermediate results in your calculator or on your computer and delay any rounding until the end of your calculations.

My thanks go to a number of people. Completing a work as large as this has required the understanding, assistance, patience, stamina, and tolerance of Neil and the staff at Addison Wesley, especially Rachel Reeve, who have been such a great help in completing this project. It has been an excellent exercise in cooperation and teamwork. I would also like to thank Vivien Freund and Lindsay Packer for their excellent work in checking all of the solutions in this manual. Finally, I want to express my thanks to my wife, Judy, for her support and encouragement throughout this project.

D.R.L.
Eau Claire, WI

CONTENTS

CHAPTER 1 ANSWERS

Exercises 1.1

1.1 (a) The *population* is the collection of all individuals, items, or data under consideration in a statistical study.

(b) A *sample* is that part of the population from which information is collected.

1.3 Descriptive methods are used for organizing and summarizing information and include graphs, charts, tables, averages, measures of variation, and percentiles.

1.5 This study is inferential. Data from a sample of Americans are used to make an estimate of (or an inference about) average TV-viewing time for all Americans.

1.7 This study is inferential. Data from a 10% sampling of all 1997 U.S. death certificates are used to make a estimates of (or inferences about) rates of the leading causes of death.

1.9 This study is descriptive. It is a summary of the final closing values of the Dow Jones Industrial Average at the end of December for the years 1994-1999.

1.11 (a) This statement is descriptive since it only tells what was said by those who were surveyed.

(b) Then the statement would be inferential since the data would have been used to provide an estimate of what <u>all</u> Americans would choose.

Exercises 1.3

1.13 A census is generally time consuming, costly, frequently impractical, and sometimes impossible.

1.15 The sample should be representative so that it reflects as closely as possible the relevant characteristics of the population under consideration.

1.17 Dentists form a high-income group whose incomes are not representative of the incomes of Seattle residents in general.

1.19 (a) Probability sampling uses a randomizing device to decide which members of the population will constitute the sample. This can involve tossing a coin or die, or consulting a random number table.

(b) False. It is possible for the randomizing device to randomly produce a sample which is not representative.

(c) Probability sampling eliminates unintentional bias, permits the researcher to control the chance of obtaining a nonrepresentative sample, and guarantees that the techniques of inferential statistics can be applied.

1.21 Simple random sampling.

1.23 (a) GLS, GLA, GLT, GSA, GST, GAT, LSA, LST, LAT, SAT.

(b) There are 10 samples, each of size three. Each sample has a one in ten chance of being selected. Thus, the probability that a sample of three salaries is the first sample on the list presented in part (a) is 1/10. The same is true for the second sample and for the tenth sample.

1.25 (a)

E,M,P,L	E,M,L,F	E,P,B,F	M,P,B,F
E,M,P,B	E,M,B,F	E,L,B,F	M,L,B,F
E,M,P,F	E,P,L,B	M,P,L,B	P,L,B,F
E,M,L,B	E,P,L,F	M,P,L,F	

(b) One procedure for obtaining a random sample of four representatives from the six is to write the initials of the representatives on six separate pieces of paper, place the six slips of paper into a box, and then, while blindfolded, pick four of the slips of paper. Or, number the representatives 1-6, and use a table of random numbers or a random-number generator to select four different numbers between 1 and 6.

(c) P(E, F, L, and B) = 1/15; P(P, B, M and F) = 1/15

1.27 I am using Table I to obtain a list of 10 random numbers between 1 and 500 as follows. First, I pick a random starting point by closing my eyes and putting my finger down on the table.

My finger falls on the three digits located at the intersection of line number 00 with columns 34, 35, and 36. The selected digits are 356. This is my starting point.

I now go down the table and record the three-digit numbers appearing directly beneath 356. Since I want numbers between 1 and 500 only, I throw out numbers between 501 and 999, inclusive. I also discard the number 000.

After 356, I skip 876, record 351, skip 717, record 239, 455, 431, 008, skip 900, 721, record 259, 068, 156, skip 570, 540, 937, 989, and record 047.

I've finished recording the 10 random numbers. In summary, these are:

356	239	431	259	156
351	455	008	068	047

1.29 (a) The possible samples of size one are G L S A T

(b) There is no difference between obtaining a simple random sample of size one and selecting one official at random.

1.31 Set a range of 1 to 500 integers and have a random-number generator generate 10 numbers in this range. Match these numbers to a numbered list of the 500 firms.

<u>Exercises 1.4</u>

1.33 (a) Answers will vary, but here is the procedure: (1) Divide the population size, 500, by the sample size, 10, and round down to the nearest whole number if necessary; this gives 50. (2) Use a table of random numbers (or a similar device) to select a number at random between 1 and 50, call it k. (3) List every 50th number, starting with k, until 10 numbers are obtained; thus the first number on the required list of 10 numbers is k, the second is k+50, the third is k+100, and so forth (e.g., if k=6, then the numbers on the list are 6, 56, 106, ...).

(b) Systematic random sampling is easier.

(c) The answer may depend on the purpose of the sampling. If the purpose of sampling is not related to the size of the sales outside the U.S., systematic sampling may work. However, since the listing is a ranking by amount of sales, if the first number chosen is low (say 2), then the sample will contain firms that, on the average, have higher sales outside the U.S. than the population as a whole. If the first number is high, say 49, then the sample will contain firms that, on the average, have lower sales than the population as a whole. In either of those cases, the sample would not be representative of the population in regard to the amount of sales outside the U.S.

1.35 (a) Number the suites from 1 to 48, use a table of random numbers to randomly select three of the 48 suites, and take as the sample the 24 dormitory residents living in the three suites obtained.

(b) Probably not, since friends are more likely to have similar opinions than are strangers.

(c) There are 384 students in total. Freshman make up 1/3 of them. Sophomores make up 7/24 of them, Juniors 1/4, and Seniors 1/8. Multiplying each of these fractions by 24 yields the proportional allocation which dictates that the number of freshmen, sophomores, juniors, and seniors selected be, respectively, 8, 7, 6, and 3. Thus a stratified sample of 24 dormitory residents can be obtained as follows: Number the freshmen dormitory residents from 1 to 128 and use a table of random numbers to randomly select 8 of the 128 freshman dormitory residents; number the sophomore dormitory residents from 1 to 112 and use a table of random numbers to randomly select 7 of the 112 sophomore dormitory residents; and so forth.

1.37 From the information about the sample, we can conclude that the population of interest consists of all adults in the continental U.S. The sample size was 2010 except that for questions about politics, only registered voters were considered part of the sample. The sample size for those questions was 1,637.

The overall procedure for drawing the sample was multistage sampling (actually, three stages were used): the first stage was to randomly select 520 geographic points in the continental U.S.; then proportional sampling was used to randomly sample a number of households with telephones from each of the 520 regions in proportion to its population; finally, once each household was selected, a randomizing procedure was used to ensure that the correct numbers of adult male and female respondents were included in the sample.

The last paragraph indicates the confidence that the poll-takers had in the results of the survey, that is, that there is a 95% chance that the sample results will not differ by more than 2.2 percentage points in either direction from the true percentage that would have been obtained by surveying all adults in the actual population, or by more than 2.5 percentage points in either direction from the true percentage that would have been obtained by surveying all registered voters in the population. The last sentence says that smaller samples have a larger "margin of error," an explanation for the difference in the maximum error for all adults and the maximum error for registered voters.

Exercises 1.5

1.39 Observational studies can reveal only association, whereas designed experiments can help establish cause and effect.

1.41 Here is one of several methods that could be used: Number the women from 1 to 4753; use a table of random numbers or a random-number generator to obtain 2376 different numbers between 1 and 4753; the 2376 women with those numbers are in one group, the remaining 2377 women are in the other group.

1.43 Designed experiment. The researchers did not simply observe the two groups of children, but instead randomly assigned children to receive the Salk vaccine or to receive a placebo.

1.45 (a) Experimental units are the individuals or items on which the experiment is performed.

(b) When the experimental units are humans, we call them subjects.

1.47 (a) Experimental units: the twenty flashlights

(b) Response variable: battery lifetime in a flashlight

(c) Factors: one factor - battery brand

(d) Levels of the factor: the four brands of batteries

(e) Treatments: the four brands of batteries

1.49 (a) Experimental units: the product being sold

(b) Response variable: the number of units of the product sold

(c) Factors: two factors - display type and pricing scheme

(d) Levels of each factor: the three different types of display and the three different pricing schemes

(e) Treatments: the nine different combinations of display type and price which result from testing each of the three pricing schemes with each of the three display types

1.51 This is a completely randomized design since all the flashlights were randomly assigned to the different battery brands.

1.53 Double-blinding guards against bias, both in the evaluations and in the responses. In the Salk-vaccine experiment, double-blinding ensured that a doctor's evaluation would not be influenced by knowing which treatment (vaccine or placebo) a patient received; it also ensured that a patient's response to the treatment from would not be influenced by knowing which treatment he or she received.

1.55 Minitab is a comprehensive statistical software package which contains features used for random sampling. Using Minitab, we wish to randomly select 2376 different numbers from the set of numbers 1 to 4753. To do this,

- From the pull-down menu, choose **Calc ▶ Make Patterned Data ▶ Simple Set of Numbers...**
- Type NUMBERS in the **Store patterned data in** text box.
- Select the **Patterned sequence** option button
- Click in the **From first value** text box and type 1
- Click in the **To last value** text box and type 4753
- Click **OK**

- Now choose **Calc ▶ Random Data ▶ Sample From Columns...**
- Type 2376 in the small text box after **Sample**
- Click in the **Sample ☐ rows from column(s)** text box and specify NUMBERS
- Click in the **Store Samples in** text box and type SRS
- Click **OK**

To print the numbers now in the SRS column, we choose **Manip ▶ Display Data...**, specify SRS in the **Columns, Constants, and Matrices to display** text box, and then click **OK**. Results are shown in the Session window and can be printed from there.

REVIEW TEST FOR CHAPTER 1

1. Student exercise.

2. Descriptive statistics are used to display and summarize the data to be used in an inferential study. Preliminary descriptive analysis of a sample often reveals features of the data that determine the choice of the appropriate inferential analysis procedure.

3. Descriptive study. The scores are merely reported.

4. Descriptive study. The paragraph describes the results of a survey of thousands of students.

5. A literature search should be made before planning and conducting a study.

6. (a) A representative sample is one which reflects as closely as possible the relevant characteristics of the population under consideration.

(b) Probability sampling involves the use of a randomizing device to determine which members of the population will make up the sample. This can involve tossing a coin or die, using a random number table, or using computer software which generates random numbers

(c) Simple random sampling is a sampling procedure in which all possible samples of a given size are equally likely to be the actual sample selected.

7. Since Yale is a highly selective and expensive institution, it is very unlikely that the students at Yale are representative of the population of all college students or that their parents are representative of the population of parents of all college students.

8. (a) This method does not involve probability sampling. No randomizing device is being used and people who do not visit the campus have no chance of being included in the sample.

(b) The dart throwing is a randomizing device which makes all samples of size 20 equally likely. This is probability sampling.

9. (a)

SW,AA,DL	SW,AA,US	SW,AA,AK	SW,DL,US	SW,DL,AK
SW,US,AK	AA,DL,US	AA,DL,AK	AA,US,AK	DL,US,AK

(b) Since each of the 10 samples of size three is equally likely, there is a 1/10 chance that the sample chosen is the first sample in the list, a 1/10 chance that it is the second sample in the list, and a 1/10 chance that it is the tenth sample in the list.

10. (a) Table I can be employed to obtain a sample of 15 different random numbers between 1 and 100 as follows. First, I pick a random starting point by closing my eyes and putting my finger down on the table.

My finger falls on three digits located at the intersection of a line with three columns. (Notice that the first column of digits is labeled "00" rather than "01"). This is my starting point.

I now go down the table and record all three-digit numbers appearing directly beneath the first three-digit number which are between 001 and 100 inclusive. I throw out numbers between 101 and 999, inclusive. I also discard the number 0000. When the bottom of the column is reached, I move over to the right to the next sequence of three digits and work my way back up the table. Continue in this manner. When 10 distinct three-digit numbers have been recorded, the sample is complete.

(b) Starting in row 10, columns 7-9, we skip 484, 797, record 082, skip 586, 653, 452, 552, 155, record 008, skip 765, move to the right and record 016, skip 534, 593, 964, 667, 452, 432, 594, 950, 670, record 001, skip 581, 577, 408, 948, 807, 862, 407, record 047, skip 977, move to the right, skip 422 and all of the rest of the numbers in that column, move to the right, skip 732, 192, record 094, skip 615 and all of the rest of the numbers in that column, move to the right, record 097, skip 673, record 074, skip 469, 822, record 052, skip 397, 468, 741, 566, 470, record 076, 098, skip 883, 378, 154, 102, record 003, skip 802, 841, move to the right, skip 243, 198, 411, record 089, skip 701, 305, 638, 654, record 041, skip 753, 790, record 063.

The final list of numbers is

082, 008, 016, 001, 047, 094, 097, 074, 052, 076, 098, 003, 089, 041, 063.

11. (a) The procedure for systematic random sampling is as follows: first we divide the population size by the sample size and round the result down to the next integer, say k. Then we select one random number, say r, between 1 and k inclusive. That number will be the first member of the sample. The remaining members of sample will be those numbered r+k, r+2k, r+3k, ... until a sample of the desired size n has been chosen. Systematic sampling will yield results similar to simple random sampling as long as there is nothing systematic about the way the members of the population were assigned their numbers.

(b) In cluster sampling, clusters of the population (such as blocks, precincts, wards, etc.) are chosen at random from all such possible clusters. Then every member of the population belonging to the chosen clusters is sampled. This method of sampling is particularly convenient when members of the population are widely scattered and is appropriate when the members of each cluster are representative of the entire population.

(c) In stratified random sampling with proportional allocation, the population is first divided into subpopulations, called strata, and simple random sampling is done within each stratum. Proportional allocation means that the size of the sample from each stratum is proportional to the size of the stratum.

12. (a) Answers will vary, but here is the procedure: (1) Divide the population size, 100, by the sample size 15, and round down to the nearest whole number; this gives 6. (2) Use a table of random numbers (or a similar device) to select a number between 1 and 6, call it k. (3) List every 6th number, starting with k, until 15 numbers are obtained; thus the first number on the required list of 15 numbers is k, the second is $k+6$, the third is $k+12$, and so forth (e.g., if $k=4$, then the numbers on the list are 4, 10, 16, ...).

(b) Yes, unless for some reason there is some kind of trend or a cyclical pattern in the listing of the athletes.

13. (a) The number of full professors should be (205/820) x 40 = 10. Similarly, proportional allocation dictates that 16 associate professors, 12 assistant professors, and 2 instructors be selected.

(b) The procedure is as follows: Number the full professors from 1 to 205, and use Table I to randomly select 10 of the 205 full professors; number the associate professors from 1 to 328, and use Table I to randomly select 16 of the 328 associate professors; and so on.

14. The statement under the vote is a disclaimer as to the validity of the survey. Since the vote reflects only the responses of volunteers who chose to vote, it cannot be regarded as representative of the public in general, some of whom do not use the Internet, nor as representative of Internet users since the sample was not chosen at random from either group.

15. (a) In an observational study, researchers simply observe characteristics and take measurements. In a designed experiment, researchers impose treatments and controls and *then* observe characteristics and take measurements.

(b) An observational study can only reveal associations between variables, whereas a designed experiment can help to establish cause and effect relationships.

16. This is an observational study. The researchers at the University of Michigan simply observed the poverty status and IQs of the children.

17. (a) This is a designed experiment.

(b) The treatment group consists of the 158 patients who took AVONEX. The control group consists of the 143 patients who were given a placebo. The treatments were the AVONEX and the placebo.

18. The three basic principles of experimental design are control, randomization, and replication. Control refers to methods for controlling factors other than those of primary interest. Randomization means randomly dividing the subjects into groups in order to avoid unintentional selection bias in constituting the groups. Replication means using enough experimental units or subjects so that groups resemble each other closely and so that there is a good chance of detecting differences among the treatments when such differences actually exist.

19. (a) Experiment units: the doughnuts

(b) Response variable: amount of fat absorbed

(c) Factor(s): fat type

(d) Levels of the factor: four types of fat

(e) Treatments: four types of fat

20. (a) Experiment units: tomato plants

(b) Response variable: yield of tomatoes

(c) Factor(s): tomato variety and density of plants

(d) Levels of each factor: These are not given, but tomato varieties tested would be the levels of variety and the different densities of plants would be the levels of density.

(e) Treatments: Each treatment would be one of the combinations of a tomato and a given plant density.

21. (a) This is a completely randomized design since all 24 cars were randomly assigned to the 4 brands of gasoline.

(b) This is a randomized block design. The four different gasoline brands are randomly assigned to the four cars separately within each of the six car model groups. The blocks are the six groups of four identical cars.

(c) If the purpose is learn about the mileage rating of one particular car model with each of the four gasolines, then the completely randomized design is appropriate. But if the purpose is to learn about the performance of the gasolines across a variety of cars (and this seems more reasonable), then the randomized block design is more appropriate and will allow the researcher to determine the effect of car model as well as of gasoline type on the mileage obtained.

24. (a) Student Exercise. This can be done easily in Minitab or Excel. In Excel, type the expression =INT(100*RAND()+1) in cell A1. This will produce a random integer between 1 and 100 inclusive. Then copy this expression into cells A2:A15. The result will be a sample of 15 numbers. If there are duplicates, copy the expression into enough cells to yield 15 different values.

CHAPTER 2 ANSWERS

Exercises 2.1

2.1 (a) Answers may vary. Eye color and model of car are qualitative variables.

(b) Answers may vary. Number of eggs in a nest, number of cases of flu, and number of employees are discrete, quantitative variables.

(c) Weight and voltage are examples of quantitative continuous variables.

2.3 (a) Qualitative data are obtained by observing the characteristics described by a qualitative variable such as color or shape.

(b) Discrete, quantitative data are numerical data that are obtained by observing values, usually by counting, of a discrete variable whose values form a finite or countably infinite set of numbers.

(c) Continuous, quantitative data are numerical data that are obtained by observing values of a continuous variable. They are usually the result of measuring something such as temperature which can take any value in a given interval.

2.5 Of qualitative and quantitative (discrete and continuous) types of data, only qualitative involves non-numerical data.

2.7 (a) The second column consists of *quantitative, discrete* data. This column provides the ranks of the cities according to their highest temperatures.

(b) The third column consists of *quantitative, continuous* data. This column provides the highest temperature on record in each of the listed cities.

(c) The information that Phoenix is in Arizona is qualitative data since it is non-numeric.

2.9 (a) The third column consists of *quantitative, discrete* data. Although the data are presented to one decimal point, the data represent the number of albums sold in millions, which can only be whole numbers.

(b) The information that *Supernatural* was performed by Santana is qualitative data since it is non-numerical.

Exercises 2.2

2.11 One of the main reasons for grouping data is that it often makes a rather complicated set of data easier to understand.

2.13 When grouping data, the three most important guidelines in choosing the classes are: (1) the number of classes should be small enough to provide an effective summary, but large enough to display the relevant characteristics of the data; (2) each piece of data must belong to one, and only one, class; and (3) whenever feasible, all classes should have the same width.

2.15 If the two data sets have the same number of data values, either a frequency distribution or a relative-frequency distribution is suitable. If, however, the two data sets have different numbers of data values, relative-frequency distributions should be used because the total of each set of relative frequencies is 1, putting both distributions on the same basis.

2.17 In the first method for depicting classes we used the notation **a<b** to mean values that are greater than or equal to **a** and up to, but not including **b**. So, for example, 30<40 represents a range of values greater than or equal to 30, but strictly less than 40. In the alternate method, we used the notation **a-b** to indicate a class that extends from **a** to **b**, including both endpoints. For example, 30-39 is a class that includes both 30 and 39. The alternate method is especially appropriate when all of the data values are integers. If the data include values like 39.7 or 39.93, the first method is preferable since the cutpoints remain integers whereas in the alternate method, the upper limits for each class would have to be expressed in decimal form such as 39.9 or 39.99.

2.19 When grouping data using classes that each represent a single possible numerical value, the midpoint of any given class would be the same as the value for that class. Thus listing the midpoints would be redundant.

2.21 The first class is 52<54. Since all classes are to be of equal width 2, the classes are presented in column 1. The last class is 74<76, since the largest data value is 75.3. Having established the classes, we tally the speed figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 35, results in each class's relative frequency. The relative frequencies are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoints which are presented in column 4.

Speed (MPH)	Frequency	Relative Frequency	Midpoint
52<54	2	0.057	53
54<56	5	0.143	55
56<58	6	0.171	57
58<60	8	0.229	59
60<62	7	0.200	61
62<64	3	0.086	63
64<66	2	0.057	65
66<68	1	0.029	67
68<70	0	0.000	69
70<72	0	0.000	71
72<74	0	0.000	73
74<76	1	0.029	75
	35	1.001	

Note that the relative frequencies sum to 1.01, not 1.00, due to round-off errors in the individual relative frequencies.

2.23 The first class is 52-53.9. Since all classes are to be of equal width, the second class has limits of 54 and 55.9. The classes are presented in column 1. The last class is 74-75.9 since the largest data value is 75.3. Having established the classes, we tally the speed figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, 35, results in each class's relative frequency which is presented in column 3. By averaging the lower limit for each class with the upper limit of the same class, we arrive at the class mark for each class. The class marks are presented in column 4.

Speed (MPH)	Frequency	Relative Frequency	Class Mark
52-53.9	2	0.057	52.95
54-55.9	5	0.143	54.95
56-57.9	6	0.171	56.95
58-59.9	8	0.229	58.95
60-61.9	7	0.200	60.95
62-63.9	3	0.086	62.95
64-65.9	2	0.057	64.95
66-67.9	1	0.029	66.95
68-69.9	0	0.000	68.95
70-71.9	0	0.000	70.95
72-73.9	0	0.000	72.95
74-75.9	1	0.029	74.95
	35	1.001	

Note that the relative frequencies sum to 1.01, not 1.00, due to round-off errors in the individual relative frequencies.

2.25 Since the data values range from 3 to 12, we could construct a table with classes based on a single value or on two values. We will choose classes with a single value because one of the classes based on two values would have contained almost half of the data. The resulting table is shown below.

Number of Pups	Frequency	Relative Frequency
3	2	0.0250
4	5	0.0625
5	10	0.1250
6	11	0.1375
7	17	0.2125
8	17	0.2125
9	11	0.1375
10	4	0.0500
11	2	0.0250
12	1	0.0125
	80	1.0000

2.27

Network	Frequency	Relative Frequency
ABC	5	0.25
CBS	8	0.40
NBC	7	0.35
	20	1.00

2.29 (a) The first class is 1<2. Since all classes are to be of equal width 1, the second class is 2<3. The classes are presented in column 1. Having established the classes, we tally the volume figures into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of observations, which is 30, results in each class's relative frequency. The relative frequencies are presented in column 3. By averaging the lower and upper class cutpoints for each class, we arrive at the class midpoint for each class. The class midpoints are presented in column 4.

Volume (100sh)	Frequency	Relative Frequency	Midpoint
1<2	4	0.13	1.5
2<3	4	0.13	2.5
3<4	2	0.07	3.5
4<5	6	0.20	4.5
5<6	3	0.10	5.5
6<7	1	0.03	6.5
7<8	2	0.07	7.5
8<9	3	0.10	8.5
9<10	1	0.03	9.5
10 & Over	4	0.13	
	30	0.99	

Note that the relative frequencies sum to 0.99, not 1.00, due to round-off errors in the individual relative frequencies.

(b) Since the last class has no upper cutpoint, the midpoint cannot be computed.

2.31 (a) The classes are presented in column 1. With the classes established, we then tally the exam scores into their respective classes. These results are presented in column 2, which lists the frequencies. Dividing each frequency by the total number of exam scores, which is 20, results in each class's relative frequency. The relative frequencies are presented in column 3. By averaging the lower and upper cutpoints for each class, we arrive at the class mark for each class. The class marks are presented in column 4.

Score	Frequency	Relative Frequency	Class Mark
30-39	2	0.10	34.5
40-49	0	0.00	44.5
50-59	0	0.00	54.5
60-69	3	0.15	64.5
70-79	3	0.15	74.5
80-89	8	0.40	84.5
90-100	4	0.20	95.0
	20	1.00	

(b) The first six classes have width 10; the seventh class has width 11.

(c) Answers will vary, but one choice is to keep the first six classes the same and make the next two classes 90-99 and 100-109.

2.33 In Minitab, place the cheetah speed data in a column named SPEED and put the *WeissStats* CD in the CD drive. Assuming that the CD drive is drive D, then type in Minitab's session window after the MTB> prompt the command %D:\IS6\Minitab\Macro\group.mac 'SPEED' and press the **ENTER** key. We are given three options for specifying the classes. Since we want the first class to have lower cutpoint 52 and a class width of 2, we select the third option (3) by entering 3 after the DATA> prompt, press the **ENTER** key, and then type 52 2 when prompted to enter the cutpoint and class width of the first class. Press the **ENTER** key again. The resulting output is

Grouped-data table for SPEED N = 35

Row	LowerCut	UpperCut	Freq	RelFreq	Midpoint
1	52	54	2	0.057	53
2	54	56	5	0.143	55
3	56	58	6	0.171	57
4	58	60	8	0.229	59
5	60	62	7	0.200	61
6	62	64	3	0.086	63
7	64	66	2	0.057	65
8	66	68	1	0.029	67
9	68	70	0	0.000	69
10	70	72	0	0.000	71
11	72	74	0	0.000	73
12	74	76	1	0.029	75

2.35 In Minitab, with the data from the Network column in a column named NETWORK,

- Choose **Stat ▶ Tables ▶ Tally...**
- Click in the **Variables** text box and select NETWORK
- Click in the **Counts** and **Percents** boxes under **Display**
- Click **OK**. The results are

NETWORK	Count	Percent
ABC	5	25.00
CBS	8	40.00
NBC	7	35.00
N=	20	

Exercises 2.3

2.37 A frequency histogram shows the actual frequencies on the vertical axis whereas the relative frequency histogram always shows proportions (between 0 and 1) or percentages (between 0 and 100) on the vertical axis.

2.39 Since a bar graph is used for qualitative data, we separate the bars from each other to emphasize that there is no numerical scale and no special ordering of the classes; if the bars were to touch, some viewers might infer an ordering and common values for adjacent bars.

2.41 (a) Each rectangle in the frequency histogram would have a height equal to the number of dots in the dot diagram.

(b) If the classes for the histogram were based on multiple values, there would not be one rectangle corresponding to each column of dots (there would be fewer rectangles than columns of dots). The height of a given rectangle would be equal to the total number of dots between its cutpoints. If the classes were constructed so that only a few columns of dots corresponded to each rectangle, the general shape of the distribution should remain the same even though the details may differ.

2.43 (a) The frequency histogram in Figure (a) is constructed using the frequency distribution presented in this exercise; i.e., columns 1 and 2. The lower class limits of column 1 are used to label the horizontal axis of the frequency histogram. Suitable candidates for vertical-axis units in the frequency histogram are the integers 0 through 8, since these are representative of the magnitude and spread of the frequencies presented in column 2. The height of each bar in the frequency histogram matches the respective frequency in column 2.

(b) The relative-frequency histogram in Figure (b) is constructed using the relative-frequency distribution presented in this exercise; i.e., columns 1 and 3. It has the same horizontal axis as the frequency histogram. We notice that the relative frequencies presented in column