STUDENT'S SOLUTIONS MANUAL

MILTON LOYER

to accompany

ELEMENTARY STATISTICS

EIGHTH EDITION

Mario F. Triola

STUDENT'S SOLUTIONS MANUAL

Milton Loyer
Penn State University

to accompany

Elementary Statistics
Eighth Edition

Mario F. Triola

Addison Wesley Longman

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

Reproduced by Addison-Wesley Publishing Company Inc. from camera-ready copy supplied by the authors.

ISBN 0-201-70465-X

3 4 5 6 7 8 9 10 PHTH 03 02

PREFACE

This manual contains the solutions to the odd-numbered exercises for each section of the textbook Elementary Statistics, Eighth Edition, by Mario Triola, and the solutions for all end-of-chapter review and cumulative review exercises of that text. In the worked problems, intermediate steps are provided for the calculations. When appropriate, additional hints and comments are included and prefaced by NOTE.

Many statistical problems are best solved using particular formats. Recognizing and following these patterns promotes understanding and develops the capacity to apply the concepts to other problems. This manual identifies and employs such formats whenever practicable.

For best results, read the text carefully before attempting the exercises, and attempt the exercises before consulting the solutions. This manual has been prepared to provide a check and extra insights for exercises that have already been completed and to provide guidance for solving exercises that have already been attempted but have not been successfully completed.

I would like to thank Mario Triola for writing an excellent elementary statistics book and for inviting me to prepare this solutions manual.

TABLE OF CONTENTS

Chapter 1

Introduction to Statistics

1-2 The Nature of Data

1. Statistic, since 20.7 refers to the selected sample.

3. Parameter, since 30 refers to the entire population.

5. Discrete, since the number of absent students must be an integer.

7. Discrete, since the number owning answering machines must be an integer.

9. Ratio, since differences are meaningful and zero height has a natural meaning.

11. Interval, since differences are meaningful but ratios are not. Refer to exercise #19.

13. Interval, since differences are meaningful but ratios are not. Years are not data at the ratio level of measurement because the year zero has been arbitrarily assigned so that the year 0 does not indicate the absence of time. The year 1900, for example, does not represent twice as much time as 950 -- and the ratio would be different using the Chinese or Jewish numerical representations for those years. Since the time difference between 1900 and 1920 is the same as the time difference between 1920 and 1940, however, years are data at the interval level of measurement.

15. Ordinal, since the ratings give relative position in a hierarchy.

17. Ratio, since differences are meaningful and zero ounces has a natural meaning.

19. Temperature ratios are not meaningful because a temperature of 0° does not represent the absence of temperature in the same sense that $0 represents the absence of money. The zero temperature in the example (whether Fahrenheit or Centigrade) was determined by a criterion other than "the absence of temperature."

1-3 Uses and Abuses of Statistics

1. Because the 186,000 respondents were self-selected and not randomly chosen, they are not necessarily representative of the general population and provide no usable information about the general population. In addition, the respondents were self-selected from a particular portion of the general population -- persons watching "Nightline" and able to spend the time and money to respond.

3. a. 500 + (.05)(500) = 500 + 25 = 525
   b. 525 - (.05)(500) = 525 - 26.25 = 498.75. No, because the 5% decrease is based on a larger amount than was the previous year's 5% increase.

5. That healthier babies are born to mothers who eat lobsters doesn't mean that eating lobster caused the babies to be healthier. Mothers who eat lobster are probably more affluent than the general population and would tend to eat better, be more knowledgeable about proper pre-natal care, have better health care, etc.

7. Motorcyclists that died in crashes in which helmets may have saved their lives could not be present to testify.

9. There are several possible answers. (1) Since tallness is perceived to be a favorable attribute, people tend to overstate their heights; at the very least, people would tend to round to the next highest inch and not to the nearest inch. (2) Many people do not really accurately know their height. (3) Because Americans tend to express height in feet and inches, errors might occur either in converting heights to all inches or in misstatements like 52" for 5'2". (4) Because many cultures express height in centimeters, some people might not know or be able to readily calculate their heights in inches.

11. No. Since the second 5% price cut would be based on a lower price, two consecutive 5% price cuts yield a smaller price reduction than a single 10% price cut. Mathematically, the two consecutive 5% cuts yield a reduction of .05x + .05(x-.05x) = .0975x, or a 9.75% price cut.

13. Assuming that each of the 20 individual subjects is ultimately counted as a success or not (i.e., that there are no "dropouts" or "partial successes"), the success rates in fraction form must be one of 0/20, 1/20, 2/20,..., 19/20, 20/20. In percentages, these rates are multiples of 5 (0%, 5%, 10%,..., 95%, 100%), and values such as 53% and 58% are not mathematical possibilities.

15. a. Since 100% is the totality of whatever is being measured, removing 100% of some quantity means that none of it is left.
    b. Reducing plaque by over 300% would mean removing three times as much plaque as is there, and then removing even more!

1-4 Design of Experiments

1. Observational study, since specific characteristics are measured on unmodified subjects.

3. Experiment, since the effect of an applied treatment is measured.

5. Random, since each 212 area code telephone number has an equal chance of being selected. But this is a really a complex situation, as indicated by the following NOTES.

   NOTE 1: This ignores the fact that some residences may have more than one phone number. A residence with two different phone numbers (e.g., one for the parents and one for the teenagers) has twice the chance of being selected as does a residence with a single phone number.

   NOTE 2: The scenario stated the organization sought to poll "residents" with the 212 area code. If the organization polls all residents at each selected number, this is cluster sampling. If the organization polls one resident at each selected number, the sample is not a random sample of "residents" because a resident living alone and having his own phone number has a higher chance of being selected than a resident living with others (e.g., in a family) and sharing a common phone number.

   NOTE 3. The poll will not include residents in the 212 area code who do not have such phone numbers. This is not a problem if the intended population is phone customers (e.g., for a poll of satisfaction with phone service), but it is if the intended population is general residents (e.g., for a poll of satisfaction with garbage service).

7. Convenience, since the sample was simply those who happen to pass by.

9. Stratified, since the population of interest (assumed to be all car owners) was divided into 5 subpopulations from which the actual sampling was done.

11. Cluster, since the population of interest (assumed to be all students at The College of Newport) was divided into classes which were randomly selected in order to interview all the students in each selected class.

    NOTE: Ideally the division into classes should place each student into one and only one class (e.g., if every student must take exactly one PE class each semester, select the PE classes at random. In practice such divisions are often made in ways that place some students in none of the classes (e.g., by selecting from all 2 pm W-W-F classes) or in more than one of the classes (e.g., by selecting from all the classes offered in the college). With careful handling, imperfect divisions do not significantly affect the results.

13. Stratified, since the population of interest (assumed to be all workers) was divided into 3 subpopulations from which the actual sampling was done.

15. Systematic, since every fifth element in the population (assumed to be all drivers passing the checkpoint during its operation) was sampled.

17. There are several possible answers. (1) Write each full-time student's name on a slip of paper, place the slips in a box, mix them thoroughly, and select 200 of them. (2) Assign each full-time a number (e.g., alphabetically), and use a table of random digits (or a calculator or a computer) to generate 200 random numbers with the appropriate numbers of digits.

19. Obtain from each college bookstore a list of the textbooks currently beings used, and compile a single master list (i.e., without duplications). Number the textbooks on the master list, and use a table of random digits (or a calculator or a computer) to generate 200 numbers with the appropriate numbers of digits.

21. a. Open questions elicit the respondent's true feelings without putting words or ideas into his mind. In addition, open questions might produce responses the pollster failed to consider. Unfortunately open questions sometimes produce responses that are rambling, unintelligible or not relevant.
    b. Closed questions help to focus the respondent and prevent misinterpretation of the question. Sometimes, however, closed questions reflect only the wording and opinions of the pollster and do not allow respondents to express legitimate alternatives.
    c. Closed questions are easier to analyze because the pollster can control the number of possible responses to each question and can word the responses to establish relationships between the responses and with other questions.

23. Confounding occurs when the researcher is not able to determine which factor (often one planned and one unplanned) produced an observed effect. If a restaurant tries adding an evening buffet for one week and it is the same week a nearby theater happens to show a real blockbuster that attracts unusual crowds to the neighborhood, the restaurant can not know whether its increased business is due to the new buffet or the extra traffic created by the theater.

25. a. Possibly; no. Stratified random sampling can employ either the same sample size for each stratum or different sample sizes for the various strata. It results in a random sample only when the sample size for each stratum is proportional to the size of the stratum -- i.e., the same proportion (and not the same number) of each stratum is selected for the sample. If the strata are all the same size, then use the same sample size for each; if one stratum is half the size of the others, then its sample size should be half the other sample sizes. If one stratum is half the size of the others and the same sample size is used for each of the strata, then an element in the smaller stratum has a larger chance of being selected than an element in a larger strata -- and that violates the definition of a random sample that requires that each element has the same chance of being selected. Stratified sampling can never result in a simple random sample. It guarantees that the total sample will always include elements from each of the strata, and that total samples without any elements from one of the strata can not occur -- and that violates the definition of simple random sampling that requires that each total sample has the same chance of being selected.
    b. Possibly; no. When each element in the population is in one and only cluster, cluster sampling always results in a random sample. The chance that any element is selected is the chance that its cluster is selected; since each cluster has the same chance of being selected, each element has the same chance of being selected -- and that satisfies the definition of a random sample. Cluster sampling can never result in a simple random sample. It guarantees that total samples with elements from each of the clusters can not occur -- and that violates the definition of simple random sampling that requires that each total sample has the same chance of being selected.

Review Exercises

1. a. Discrete, since the number of shares held must be an integer.
   NOTE: Even if partial shares are allowed (e.g., 5½ shares), the number of shares must be some fractional value and not any value on a continuum -- e.g., a person could not own π shares.
   b. Ratio, since differences between values are consistent and there is a natural zero.
   c. Stratified, since the set of interest (all stockholders) was divided into subpopulations (by states) from which the actual sampling was done.
   d. Statistic, since the value is determined from a sample and not the entire population.
   e. There is no unique correct answer, but the following are reasonable possibilities. (1) The proportion of stockholders holding above that certain number of shares (which would vary from company to company) that would make them "influential." The proportion of stockholders holding below that certain number of shares (which would vary from company to company) that would make them "insignificant." (3) The numbers of shares (and hence the degree of influence) held by the largest stockholders.
   f. There are several possible valid answers. (1) The results would be from a self-selected group (i.e., those who chose to respond) and not necessarily a representative group. (2) If the questionnaire did not include information on the numbers of shares owned, the views of small stockholders (who are probably less knowledgeable about business and stocks) could not be distinguished from those of large stockholders (whose views should carry more weight).

2. a. Systematic, since the selections are made at regular intervals.
   b. Convenience, since those selected were the ones who happened to attend.
   c. Cluster, since the stockholders were organized into groups (by stockbroker) and all the stockholders in the selected groups were chosen.
   d. Random, since each stockholder has the same chance of being selected.
   e. Stratified, since the stockholders were divided into subpopulations from which the actual sampling was done.

3. Let N be the total number of full-time students and n be the desired sample size.
   a. Random. Obtain a list of all N full-time students, number the students from 1 to N, select n random numbers from 1 to N, and poll each student whose number on the list is one of the random numbers selected.
   b. Systematic. Obtain a list of all N full-time students, number the students from 1 to N, let m be the largest integer less than the fraction N/n, select a random number between 1 and m, begin with the student whose number is the random number selected, and poll that student and every mth student thereafter.
   c. Convenience. Select a location (e.g., the intersection of major campus walkways) by which most of the students usually pass, and poll the first n full-time students that pass by.
   d. Stratified. Obtain a list of all N full-time students and the gender of each, divide the list by gender, and randomly select and poll n/2 students from each gender.
   e. Cluster. Obtain a list of all the classes meeting at a popular time (e.g., 10 am Monday), estimate how many of the classes would be necessary to include n students, select that many of the classes at random, and poll all of the students in each selected class.

4. a. Blinding occurs when those involved in an experiment (either as subjects or evaluators) do not know whether they are dealing with a treatment or a placebo. It might be used in this experiment by (a) not telling the subjects whether they are receiving Sleepeze or the placebo and/or (b) not telling any post-experiment interviewers or evaluators which subjects received Sleepeze and which ones received the placebo. Double-blinding occurs when neither the subjects nor the evaluators know whether they are dealing with a treatment or a placebo.
   b. The data reported will probably involve subjective assessments (e.g., "On a scale of 1 to 10, how well did it work?") that may be subconsciously influenced by whether the subject was known to have received Sleepeze or the placebo.
   c. In a completely randomized block design, subjects are assigned to the groups (in this case to receive Sleepeze or the placebo) at random.
   d. In a rigorously controlled block design, subjects are assigned to the groups (in this case to receive Sleepeze or the placebo) in such a way that the groups are similar with respect to extraneous variables that might affect the outcome. In this experiment it may be important to make certain each group has approximately the same age distribution, degree of insomnia, number of males, number users of alcohol and/or tobacco, etc.
   e. Replication involves repeating the experiment on a sample of subjects large enough to ensure that atypical responses of a few subjects will not give a distorted view of the true situation.

5. The sample is essentially a convenience sample that might not be representative of the student body. In particular, students likely to drop out may exhibit certain common characteristics (e.g., sleeping in and/or cutting classes) that would make them under-represented in the sample because they would be less likely to pass by the polling location.

6. a. Ratio, since differences are meaningful and zero milligrams of tar has a natural meaning.
   b. Ordinal, since the ratings give relative position in a hierarchy.
   c. Nominal, since the classifications only identify categories and not relative positions on a scale.
   d. Ordinal, since the scores give relative position in a hierarchy but differences are not meaningful -- i.e., the difference in intelligence between IQ's of 40 and 50 is not the same as the difference in intelligence between IQ's of 100 and 110.
   e. Ratio, since differences are meaningful and zero points scored has a natural meaning.

Cumulative Review Exercises

NOTE: Here and throughout the text intermediate mathematical steps will be shown as an aide to those who may be having difficulty with the calculations. In practice, most of the work can be done continuously on calculators and the intermediate values are unnecessary. Even when the calculations can not be done continuously, DO NOT WRITE AN INTERMEDIATE VALUE ON YOUR PAPER AND THEN RE-ENTER IT IN THE CALCULATOR. That can practice can introduce round-off errors and copying errors. Store any intermediate values in the calculator so that you can recall them with infinite accuracy and without copying errors.

1. $\dfrac{1.23 + 4.56 + 7.89}{5} + \dfrac{13.68}{5} = 4.56$

2. $\sqrt{\frac{(5-7)^2 + (12-7)^2 + (4-7)^2}{3-1}} = \sqrt{\frac{(-2)^2 + (5)^2 + (-3)^2}{2}} = \sqrt{\frac{4 + 25 + 9}{2}} = \sqrt{\frac{38}{2}} = \sqrt{19}$
$= 4.359$

3. $\frac{1.96^2 \cdot (0.4)(0.6)}{0.025^2} = \frac{3.8416 \cdot (.24)}{.000625} = 1475.174$

4. $\frac{98.20 - 98.60}{0.62/\sqrt{106}} = \frac{-.40}{.0602} = -6.642$

5. $\frac{25!}{16!9!} = \frac{25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16!}{16!9!} = \frac{25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 \cdot 19 \cdot 18 \cdot 17}{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2042975$

NOTE: This exercise should be worked using the factorial key on the calculator. The above intermediate steps illustrate the mathematical logic involved but do not represent an efficient approach to the problem.

6. $\sqrt{\frac{10(513.27) - 71.5^2}{10(9)}} = \sqrt{\frac{5132.7 - 5112.25}{90}} = \sqrt{\frac{20.45}{90}} = \sqrt{.2272} = .477$

7. $\frac{8(151{,}879) - (516.5)(2176)}{\sqrt{8(34{,}525.75) - 516.5^2}\sqrt{8(728{,}520) - 2176^2}} = \frac{1215032 - 1123904}{\sqrt{9433.75}\sqrt{1093184}} = \frac{91128}{101552.0} = .897$

8. $\frac{(183 - 137.09)^2}{137.09} + \frac{(30 - 41.68)^2}{41.68} = \frac{(45.91)^2}{137.09} + \frac{(-11.68)^2}{41.68}$

$= \frac{2107.7281}{137.09} + \frac{136.4224}{41.68}$

$= 15.375 + 3.273$

$= 18.648$

9. $0.95^{150} = 4.56\text{E-}04 = 4.56 \cdot 10^{-4} = .000456$; moving the decimal point left 4 places

NOTE: Calculators and computers vary in their representation of such numbers. This manual assumes they will be given in scientific notation as a two-decimal number between 1.00 and 9.99 inclusive followed by an indication (usually E for *exponent* of the multiplying power of ten) of how to adjust the decimal point to obtain a number in the usual notation (rounded off to three significant digits).

10. $25^8 = 1.53\text{E+}11 = 1.53 \cdot 10^{11} = 153{,}000{,}000{,}000$; moving the decimal point right 11 places

11. $52^6 = 1.98\text{E+}10 = 1.98 \cdot 10^{10} = 19{,}800{,}000{,}000$; moving the decimal point right 10 places

12. $.25^5 = 9.77\text{E-}04 = 9.77 \cdot 10^{-4} = .000977$; moving the decimal point left 4 places

Chapter 2

Describing, Exploring, and Comparing Data

2-2 Summarizing Data with Frequency Tables

1. Subtracting two consecutive lower class limits indicates that the class width is 60 - 55 = 5. Since there is a gap of 1.0 between the upper class limit of one class and the lower class limit of the next, class boundaries are determined by increasing or decreasing the appropriate class limits by (1.0)/2 = 0.5. The class boundaries and class midpoints are given in the table below.

height	class boundaries	class midpoint	frequency
55 - 59	54.5 - 59.5	57	1
60 - 64	59.5 - 64.5	62	3
65 - 69	64.5 - 69.5	67	49
70 - 74	69.5 - 74.5	72	46
75 - 79	74.5 - 79.5	77	1
			100

NOTE: Although they often contain extra decimal points and may involve consideration of how the data were obtained, class boundaries are the key to tabular and pictorial data summaries. Once the class boundaries are obtained, everything else falls into place. Here the first class width is readily seen to be 59.5 - 54.5 = 5.0 and the first midpoint is (54.5 + 59.5)/2 = 57. In this manual, class boundaries will typically be calculated first and then used to determine other values. In addition, the sum of the frequencies is an informative number used in many subsequent calculations and will be shown as an integral part of each table.

3. Since the gap between classes as presented is .01, the appropriate class limits are increased or decreased by (.01)/2 = .005 to obtain the class boundaries and the following table.

GPA	class boundaries	class midpoint	frequency
0.00 - 0.49	-0.005 - 0.495	0.245	72
0.50 - 0.99	0.495 - 0.995	0.745	23
1.00 - 1.49	0.995 - 1.495	1.245	47
1.50 - 1.99	1.495 - 1.995	1.745	135
2.00 - 2.49	1.995 - 2.495	2.245	288
2.50 - 2.99	2.495 - 2.995	2.745	276
3.00 - 3.49	2.995 - 3.495	3.245	202
3.50 - 3.99	3.495 - 3.995	3.745	97
			1140

The class width is 0.495 - (-0.005) = .50; the first midpoint is (-0.005 + 0.495)/2 = 0.245.

5. The relative frequency for each class is found by dividing its frequency by 100, the sum of the frequencies. NOTE: As before, the sum is included as an integral part of the table. For relative frequencies, this should always be 1.000 (i.e., 100%) and serves as a check for the calculations.

height	relative frequency
55 - 59	.01
60 - 64	.03
65 - 69	.49
70 - 74	.46
74 - 79	.01
	1.00

7. The relative frequency for each class is found by dividing its frequency by 1140, the sum of the frequencies. NOTE: In #5, the relative frequencies were expressed as decimals; here they are expressed as percents The choice is arbitrary.

GPA	relative frequency
0.00 - 0.49	6.32%
0.50 - 0.99	2.02%
1.00 - 1.49	4.12%
1.50 - 1.99	11.84%
2.00 - 2.49	25.26%
2.50 - 2.99	24.21%
3.00 - 3.49	17.72%
3.50 - 3.99	8.51%
	100.00%

9. The cumulative frequencies are determined by repeated addition of successive frequencies to obtain the combined number in each class and all previous classes. NOTE: Consistent with the emphasis that has been placed on class boundaries, we choose to use upper class boundaries in the "less than" column. Conceptually, heights occur on a continuum and the integer values reported are assumed to be the nearest whole number representation of the precise measure of height. An exact height of 59.7, for example, would be reported as 60 and fall in the second class. The values in the first class, therefore, are better described as being "less than 59.5" (using the upper class boundary) than as being "less than 60." This distinction becomes crucial in the construction of pictorial representations in the next section. In addition, the fact that the final cumulative frequency must equal the total number (i.e, the sum of the frequency column) serves as a check for calculations. The sum of cumulative frequencies, however, has absolutely no meaning and is not included.

height	cumulative frequency
less than 59.5	1
less than 64.5	4
less than 69.5	53
less than 74.5	99
less than 79.5	100

11. The cumulative frequencies are determined by repeated addition of successive frequencies to obtain the combined number in each class and all previous classes. NOTE: Consistent with the emphasis that has been placed on class boundaries, we choose to use upper class boundaries in the "less than" column.

GPA	cumulative frequency
less than 0.495	72
less than 0.995	95
less than 1.495	142
less than 1.995	277
less than 2.495	565
less than 2.995	841
less than 3.495	1043
less than 3.995	1140

13. Assuming that "start the first class at 0.7900 lb" refers to the first lower class limit produces the frequency table below and violates none of the guidelines for constructing frequency tables.

weight (lbs)	frequency
.7900 - .7949	1
.7950 - .7999	0
.8000 - .8049	1
.8050 - .8099	3
.8100 - .8149	4
.8150 - .8199	17
.8200 - .8249	6
.8250 - .8299	4
	36

NOTE: The class boundaries above are .78995, .79495, .79995, etc. Using 0.7900 as the first lower class boundary produces boundaries of .7900, .7950, .8000, etc. This is not acceptable, as these are possible data values. This introduces subjectivity about where to place a value that falls on the boundary and violates the guideline that each of the values must belong to only one class.

14. Assuming that "start the first class at 0.7750 lb" refers to the first lower class limit produces the frequency table below and violates the guideline that frequency tables should have between 5 and 20 classes.

weight (lbs)	frequency
.7750 - .7799	4
.7800 - .7849	13
.7850 - .7899	15
.7900 - .7949	4
	36

NOTE: That this frequency table has only 4 categories, which is usually not sufficient to give a picture of the nature of the distribution, is allowable in this context -- since the class limits employed work well with the other cola data and permit meaningful comparisons across the data sets.

15. Assuming that "start the first class at 0.8100 lb" refers to the first lower class limit produces the frequency table below.

weight (lbs)	frequency
.8100 - .8149	1
.8150 - .8199	6
.8200 - .8249	16
.8250 - .8299	8
.8300 - .8349	3
.8350 - .8399	1
.8400 - .8449	1
	36

While similar to the frequency table in exercise #13, this table differs in two ways. (1) In exercise #13 [Coke], there were 5 classes below the modal class and 2 above; in exercise #15 [Pepsi], there are 2 classes below the modal class and 4 above. (2) The weights in exercise #13 appear to be less than those in exercise #15.

16. Assuming that "start the first class at 0.7700 lb" refers to the first lower class limit produces the frequency table below.

weight (lbs)	frequency
.7700 - .7749	1
.7750 - .7799	6
.7800 - .7849	14
.7850 - .7899	13
.7900 - .7949	2
	36

While similar to the frequency table in exercise #15, this table differs in two ways. (1) In exercise #15 [regular Pepsi], there were 2 classes below the modal class and 4 above; in exercise #16 [diet Pepsi], there are 2 classes below the modal class and 4 above. In both cases, however, there are more values above the modal class than below it. (2) The weights in exercise #15 appear to be greater than those in exercise #16.

17. For 11 classes to cover data ranging from a beginning lower class limit of 0 to a maximum value of 514, the class width must be at least (514 - 0)/11 = 46.7. A convenient class width would be 50, which produces the frequency table given below.

weight (lbs)	frequency
00 - 49	6
50 - 99	10
100 - 149	10
150 - 199	7
200 - 249	8
250 - 299	2
300 - 349	4
350 - 399	3
400 - 449	3
450 - 499	0
500 - 549	1
	36

19. Assuming that "start the first class at 200 lb" refers to the first lower class limit produces the frequency table below.

weight (lbs)	frequency
200 - 219	12
220 - 239	9
240 - 259	18
260 - 279	84
280 - 299	52
	175

Yes. Since the lowest recorded weight before collapse is over 200 and most of the weights are over 260, it appears the cans will withstand a pressure that varies between 158 and 165.

20. Assuming that "start the first class at 200 lb" refers to the first lower class limit produces the following frequency table.

weight (lbs)	frequency
200 - 219	6
220 - 239	5
240 - 259	12
260 - 279	36
280 - 299	87
300 - 319	28
	174

Yes. Most of the thicker cans support a weight of 280 before collapse, and they appear to be stronger. Since the thinner cans already meet the criterion given in exercise #19, however, the added strength of the thicker cans may not be worth the added cost.

21. Assuming that "start the first class at 200 lb" refers to the first lower class limit produces the frequency table below.

weight (lbs)	frequency
200 - 219	6
220 - 239	5
240 - 259	12
260 - 279	36
280 - 299	87
300 - 319	28
320 - 339	0
340 - 359	0
360 - 379	0
380 - 399	0
400 - 419	0
420 - 439	0
440 - 459	0
460 - 479	0
480 - 499	0
500 - 519	1
	175

In general, an outlier can add several rows to a frequency table. Even though most of the added rows have frequency zero, the table tends to suggest that these are possible valid values -- thus distorting the reader's mental image of the distribution.

23. The two frequency tables are given below.

a.

height	frequency
66 - 67	4
68 - 69	3
70 - 71	10
72 - 73	10
74 - 75	0
76 - 77	1
	28

b.

height	frequency
66 - 67	6
68 - 69	4
70 - 71	4
72 - 73	4
74 - 75	4
76 - 77	6
	28

Data set (b) appears to be the phony data for two reasons. (1) The frequencies in set (b) follow a regular pattern unlikely to be achieved by chance, while the frequencies in set (a) follow the type of irregular pattern expected by chance. (2) The pattern in (b) [heights fairly uniformly distributed with more at the extremes than near the middle] disagrees with the generally accepted pattern in (a) [many heights near the middle values and fewer at the extremes].