## REVIEW SECTION I: CHAPTERS 1-3 SOLUTIONS

1. Driver-side and passenger-side air bags are installed in all new cars to prevent serious or fatal injury in an automobile crash. However, air bags have been found to cause deaths in children and small people or people with handicaps in low-speed crashes. Consequently, in 1998 the federal government began allowing vehicle owners to request installation of an on-off switch for air bags. The table describes the reasons for requesting the installation of passenger-side on-off switches given by car owners in 1998 and 1999.

| Reason | Number of requests | RF |  |
| :--- | :--- | :--- | :---: |
| Infant | 1,852 | 0.061 |  |
| Child | 17,148 | 0.565 |  |
| Medical | 8,377 | 0.276 |  |
| Infant \& medical | 44 | 0.001 |  |
| Child \& medical | 903 | 0.030 |  |
| Infant \& child | 1,878 | 0.062 |  |
| Infant \& child \& medical | 135 | 0.004 |  |
| Total | 30,337 |  |  |

Source: National Highway Transportation Safety Administration, September 2000.
a. What type of variable, quantitative or qualitative, is summarized in the table? Give the values that the variable could assume. QUALITATIVE
b. Calculate the relative frequencies for each reason.
c. Which graph would you use to display the information in the table? PIE CHART
d. What proportion of the car owners who requested on-off air bag switches gave medical as one of the reasons? $27.6 \%+0.1 \%+3 \%+0.4 \%=31.1 \%$
2. Customer satisfaction and loyalty are valued and monitored by all world-class organizations. But are satisfied customers necessarily loyal customers? Harte-Hanks Market Research surveyed customers of department stores and banks and published the following results in American Demographics (Aug. 1999).

|  | Banks | Department <br> Stores |
| :--- | :--- | :--- |
| Totally satisfied and very loyal | $27 \%$ | $4 \%$ |
| Totally satisfied and not very loyal | $18 \%$ | $25 \%$ |
| Not totally satisfied and very loyal | $13 \%$ | $2 \%$ |
| Not totally satisfied and not very loyal | $\frac{42 \%}{100 \%}$ | $\frac{69 \%}{100 \%}$ |

Source: American Demographics,Aug. 1999.
a. Construct side-by-side relative frequency bar charts for banks and department stores.
b. Could these data have been described using pie charts? Explain. YES. TWO PIE CHARTS SIDE BY SIDE.
c. Do the data indicate that customers who are totally satisfied are very loyal? Explain. YES FOR

BANKS: 27\% ARE BOTH TOTALLY SATISFIED AND VERY LOYAL AS OPPOSED TO 18\% WHO ARE TOTALLY SATISFIED AND NOT VERY LOYAL. NO FOR DEPARTMENT STORES (4\% VS. 25\%).

3. "Made in the USA" is a claim stated in many product advertisements or on product labels. Advertisers want consumers to believe that the product is manufactured with $100 \%$ U.S. labor and materials-which is often not the case. What does "Made in the USA" mean to the typical consumer? To answer this question, a group of marketing professors conducted an experiment at a shopping mall in Muncie, Indiana (Journal of Global Business, Spring 2002). They asked every fourth adult entrant to the mall to participate in the study. A total of 106 shoppers agreed to answer the question, "Made in the USA' means what percentage of U.S. labor and materials?" The responses of the 106 shoppers are summarized in the following table.

| Response to "Made in the <br> USA" | Number of shoppers |
| :--- | :--- |
| $100 \%$ | 64 |
| 75 to $99 \%$ | 20 |
| 50 to $74 \%$ | 18 |
| Less than $50 \%$ | 4 |

Source: " 'Made in the USA': Consumer Perceptions, Deception and Policy Alternatives," Journal of Global Business, Vol. 13, No. 24, Spring 2002 (Table 3).
a. What type of data collection method was used? SURVEY
b. What type of variable, quantitative or qualitative, is measured? QUANTITATIVE
(PERCENTAGE OF LABOR AND MATERIAL IN THE PRODUCTION OF 'MADE IN USA; PRODUCTS).
4. "Reader-response cards" are used by marketers to advertise their product and obtain sales leads. These cards are placed in magazines and trade publications. Readers detach and mail in the cards to indicate their interest in the product, expecting literature or a phone call in return. How effective are these cards (called "bingo cards" in the industry) as a marketing tool? Performark, a Minneapolis business that helps companies close on sales leads, attempted to answer this question by responding to 17,000 card-advertisements placed by industrial marketers in a wide variety of trade publications over a 6 -year period. Performark kept track of how long it
took for each advertiser to respond. A summary of the response times, reported in Inc. magazine (July 1995), is given in the table.

| Advertiser's Response Time | Percentage |
| :--- | :--- |
| Never responded | 21 |
| $13-59$ days | 33 |
| $60-120$ days | 34 |
| More than 120 days | 12 |
| Total | 100 |

a. Describe the variable measured by Performark. RESPONSE TIME
b. How many of the 17,000 advertisers never responded to the sales lead? 21\% OF 17,000 = 3570
d. Advertisers typically spend at least a million dollars on a reader-response card marketing campaign. Many industrial marketers feel these "bingo cards" are not worth their expense. Will a pie chart provide information to support this contention? Explain why or why not. If not, what information can be gleaned from the pie chart to help potential "bingo card" campaigns? NO, IT WILL ONLY PROVIDE INFORMATION ON THE SPEED OF RESPONSE
5. It's not uncommon for hearing aids to malfunction and cancel the desired signal. IEEE Transactions on Speech and Audio Processing (May 1995) reported on a new audio processing system designed to limit the amount of signal cancellation that may occur. The system utilizes a mathematical equation that involves a variable, $V$, called a sufficient norm constraint. A histogram for realizations of $V$, produced using simulation, is shown below.
a. Estimate the percentage of realizations of $V$ with values ranging from .425 to .675 .
13.5+11.5+8+6+5.5=44.5\% (ESTIMATED)
b. Cancellation of the desired signal is limited by selecting a norm constraint $V$. Find the value of $V$ for a company that wants to market the new hearing a so that only $10 \%$ of the realizations have value below the selected level. ESTIMATED $\mathrm{n}=102$ (SUM OF ALL THE COLUMNS IN THE HISTOGRAM). 10\% IS 10.2 SO LOOKING FOR A VALUE OF V SUCH THAT 10.2 ARE BELOW IT. IT WOULD BE 0.325 (THE FIRST TWO CATEGORIES ARE ~2+8=10)


Source: Hoffman, M. W., and Buckley, K. M. "Robust Time-Domair Processing of Broadband Microphone Array Data." IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 3, May 1995, p. 199 (Figure 4). © 1995 IEEE.
6. Explain how the relationship between the mean and median provides information about the symmetry or skewness of the data's distribution.


IF MEAN>MEDIAN $\rightarrow$ RIGHT SKEW
IF MEAN<MEDIAN $\rightarrow$ LEFT SKEW
IF MEAN=MEDIAN $\rightarrow$ SYMMETRIC


#### Abstract

7. In order to become a certified public accountant (CPA), you must pass the Uniform CPA Exam. Many states require a minimum of 150 semester hours of college education before a candidate can sit for the CPA Exam. However, traditionally, colleges only require 128 semester hours for an undergraduate degree. A study of whether the "extra" 22 hours of college credit is warranted for CPA candidates was published in the Journal of Accounting and Public Policy (Spring 2002). For one aspect of the study, researchers sampled over 100,000 first-time candidates for the CPA exam and recorded the total semester hours of college credit for each candidate. The mean and median for the data set were 141.31 and 140 hours, respectively. Interpret these values. Make a statement about the type of skewness, if any, that exists in the distribution of total semester hours. THE MEAN IS ONLY SLIGHLY HIGHER THAN THE MEDIAN AND CONSIDERING THE SAMPLE SIZE THIS DISTIBUTION IS PROBABLY SYMMETRIC.


8. Would you expect the data sets described below to possess relative frequency distributions that are symmetric, skewed to the right, or skewed to the left? Explain.
a. The salaries of all persons employed by a large university RIGHT-SKEWED (THERE MIGHT

BE SOME VERY HIGH PAYING POSITIONS)
b. The grades on an easy test LEFT-SKEWED (MOSTLY HIGH GRADE)
c. The grades on a difficult test RIGHT-SKEWED (MOSTLY LOW GRADES)
d. The amounts of time students in your class studied last week SYMMETRIC (RANDOM GROUP OF STUDENTS)
e. The ages of automobiles on a used-car lot SYMMETRIC (RANDOM GROUP OF CARS) f. The amounts of time spent by students on a difficult examination (maximum time is 50 minutes)

LEFT-SKEWED (TAKES A LONG TIME)
9. The salaries of superstar professional athletes receive much attention in the media. The multimillion-dollar long-term contract is now commonplace among this elite group. Nevertheless, rarely does a season pass without negotiations between one or more of the players' associations and team owners for additional salary and fringe benefits for all players in their particular sports.
a. If a players' association wanted to support its argument for higher "average" salaries, which measure of central tendency do you think it should use? Why? MEDIAN. THE DISTRIBUTION IS PROBABLY RIGHT SKEWED, PULLED UP BY SOME HIGHLY PAID STAR PLAYERS. IN THIS CASE THE MEAN IS HIGHER THAN THE MEDIAN.
b. To refute the argument, which measure of central tendency should the owners apply to the players' salaries? Why? MEAN. SAME EXPLANATION AS ABOVE.
10. Calculate the variance and standard deviation for samples where:
a. $\mathrm{n}=10 ; \sum \mathrm{x} 2=84 ; \sum \mathrm{x}=20 ;$ VARIANCE $=80$, STD DEV $=8.944$
b. $n=40 ; \sum x 2=380 ; \sum x=100 ;$ VARIANCE $=373.75$, STD DEV $=19.332$
c. $n=20 ; \sum x 2=18 ; \sum x=17 ;$ VARIANCE $=17.278$, STD DEV $=4.157$
11. Compute the average, standard deviation, and variance for each of the following data sets. If appropriate, specify the units in which your answer is expressed.
a. $3,1,10,10,4$
b. 8 feet, 10 feet, 32 feet, 5 feet
c. $-1,-4,-3,1,-4,-4$
d. $1 / 5$ ounce, $1 / 5$ ounce, $1 / 5$ ounce, $2 / 5$ ounce, $1 / 5$ ounce, $4 / 5$ ounce

| DATA SET | AVERAGE | STD DEV | VARIANCE |
| :--- | :--- | :--- | :--- |
| 1 | 5.6 | 17.3 | 4.159 |
| 2 | 13.75 feet | 152.256 sqft | 12.339 feet |
| 3 | -2.5 | 4.3 | 2.074 |
| 4 | 0.33 ounce | 0.0587 squared ounces | 0.242 ounce |

12. Using only integers between 0 and 10, construct two data sets with at least 10 observations each so that the two sets have the same mean but different variances.

|  | MEAN | VARIANCE |
| :--- | :--- | :--- |
| $5,5,5,5,5,5,5,5,5,5$ | 5 | 0 |
| $4,4,4,4,4,6,6,6,6,6$ | 5 | 1.111 |

13. A widely used technique for estimating the length of time it takes workers to produce a product is the time study. In a time study, the task to be studied is divided into measurable parts and each is timed with a stopwatch or filmed for later analysis. For each worker, this process is repeated many times for each subtask. Then the average and standard deviation of the time required to complete each subtask are computed for each worker. A worker's overall time to complete the task under study is then determined by adding his or her subtask-time averages (Gaither, Production and Operations Management, 1996).The data (in minutes) given in the table are the result of a time study of a production operation involving two subtasks.

MGTS1000 - review section I; solutions

TIMESTUDY

|  | Worker A |  | Worker B |  |
| :--- | :--- | :--- | :--- | :--- |
| Repetition | Subtask1 | Subtask2 | Subtask1 | Subtask2 |
| 1 | 30 | 2 | 31 | 7 |
| 2 | 28 | 4 | 30 | 2 |
| 3 | 31 | 3 | 32 | 6 |
| 4 | 38 | 3 | 30 | 5 |
| 5 | 25 | 2 | 29 | 4 |
| 6 | 29 | 4 | 30 | 1 |
| 7 | 30 | 3 | 31 | 4 |
| AVERAGE | $\mathbf{3 0 . 1 4 3}$ | $\mathbf{3}$ | $\mathbf{3 0 . 4 2 9}$ | $\mathbf{4 . 1 4 3}$ |
| STD DEV | $\mathbf{3 . 9 7 6}$ | $\mathbf{0 . 8 1 6}$ | $\mathbf{0 . 9 7 6}$ | $\mathbf{2 . 1 1 6}$ |

a. Find the overall time it took each worker to complete the manufacturing operation under study.

WORKER 1: 30.143+3=33.143
WORKER 2: 30.429+4.143=34.572
b. For each worker, find the standard deviation of the seven times for subtask 1.
c. In the context of this problem, what are the standard deviations you computed in part b measuring? THE DEVIANCE IN COMPLETION TIMES OF SUBTASK 1
d. Repeat part b for subtask 2.
e. If you could choose workers similar to $A$ or workers similar to $B$ to perform subtasks 1 and 2, which type would you assign to each subtask? Explain your decisions on the basis of your answers to parts a-d. WORKER 1 HAS A SLIGHLTY LOWER AVERAGE FOR SUBTASK 1 BUT A MUCH HIGHER STD DEV. I'LL THEREFORE CHOOSE WORKER B FOR SUBTASK 1. WORKER A HAS BOTH A LOWER AVERAGE AND LOWER STD DEV FOR SUBTASK 2.
14. Consider the following sample of five measurements: $2,1,1,0,3$.
a. Calculate the range, $\mathrm{s}^{2}$ and $s$. RANGE $=3, \mathrm{~S}^{2}=1.3, \mathrm{~S}=1.14$
b. Add 3 to each measurement and repeat part a. RANGE=3, $S^{2}=1.3, S=1.14$
c. Subtract 4 from each measurement and repeat part a. RANGE $=3, S^{2}=1.3, S=1.14$
d. Considering your answers to parts $\mathrm{a}, \mathrm{b}$, and c , what seems to be the effect on the variability of a data set by adding the same number to or subtracting the same number from each measurement? NO EFFECT
15. The output from a statistical software package indicates that the mean and standard deviation of a data set consisting of 200 measurements are $\$ 1,500$ and $\$ 300$, respectively.
a. What are the units of measurement of the variable of interest? Based on the units, what type of data is this: quantitative or qualitative? DOLLARS; QUANTITATIVE
16. For each day of last year, the number of vehicles passing through a certain intersection was recorded by a city engineer. One objective of this study was to determine the percentage of days that more than 425 vehicles used the intersection. Suppose the mean for the data was 375 vehicles per day and the standard deviation was 25 vehicles.
a. What can you say about the percentage of days that more than 425 vehicles used the intersection? Assume you know that the relative frequency distribution for the data is mound shaped? 425 IS +2 STANDARD DEVIATIONS ABOVE THE MEAN ( $375+2 * 25=425$ ). WE KNOW THAT 95\% OF THE OBSERVATIONS FALL WITHIN $\pm 2$ STANDARD DEVIATIONS FROM THE MEAN. THEREOFRE WE CAN SAY THAT IT WAS ABOUT 2.5\%.

MGTS1000 - review section I; solutions
17. A chemical company produces a substance composed of $98 \%$ cracked corn particles and $2 \%$ zinc phosphide for use in controlling rat populations in sugarcane fields. Production must be carefully controlled to maintain the $2 \%$ zinc phosphide because too much zinc phosphide will cause damage to the sugarcane and too little will be ineffective in controlling the rat population. Records from past production indicate that the distribution of the actual percentage of zinc phosphide present in the substance is approximately mound shaped, with a mean of $2.0 \%$ and a standard deviation of $.08 \%$.
a. If the production line is operating correctly approximately what proportion of batches from a day's production will contain less than $1.84 \%$ of zinc phosphide? $\sim 2.5 \%$ (EMPIRICAL RULE) b. Suppose one batch chosen randomly actually contains $1.80 \%$ zinc phosphide. Does this indicate that there is too little zinc phosphide in today's production? Explain your reasoning. PROBABLY. IN THE ABOVE DISTRIBUTION THE CHANCES OF DRAWING 1.80\% ARE EXTREMELY LOW SINCE IT'S ALMOST 3 STD DEV FROM THE MEAN.
18. A buyer for a lumber company must decide whether to buy a piece of land containing 5,000 pine trees. If 1,000 of the trees are at least 40 feet tall, the buyer will purchase the land; otherwise, he won't. The owner of the land reports that the height of the trees has a mean of 30 feet and a standard deviation of 3 feet. Based on this information, what is the buyer's decision? 40 FEET IS MORE THAN 3 STANDARD DEVIATIONS FROM THE MEAN HEIGHT OF TREES IN THE LOT. IT'S UNLIKELY THAT AT LEAST 1000 TREES WILL BE AT LEAST 40 FEET TALL. THE BUYER WILL NOT BUY THE LOT.
19. When it is working properly, a machine that fills 25 -pound bags of flour dispenses an average of 25 pounds per fill; the standard deviation of the amount of fill is .1 pound. To monitor the performance of the machine an inspector weighs the contents of a bag coming off the machine's conveyor belt every half-hour during the day. If the contents of two consecutive bags fall more than 2 standard deviations from the mean (using the mean and standard deviation given above), the filling process is said to be out of control and the machine is shut down briefly for adjustments. The data given in the table below are the weights measured by the inspector yesterday. Assume the machine is never shut down for more than 15 minutes at a time. At what times yesterday was the process shut down for adjustment? Justify your answer.

FLOUR

| Time | Weight (pounds) | DIFFERENCE <br> FROM MEAN |
| :--- | :--- | ---: |
| 8:00A.M. | 25.10 | 0.1 |
| $8: 30$ | 25.15 | 0.15 |
| $9: 00$ | 24.81 | -0.19 |
| $9: 30$ | 24.75 | -0.25 |
| $10: 00$ | 25.00 | 0 |
| $10: 30$ | 25.05 | 0.05 |
| $11: 00$ | 25.23 | 0.23 |
| $11: 30$ | 25.25 | 0.25 |
| $12: 00$ | 25.01 | 0.01 |
| $12: 30$ P.M. | 25.06 | 0.06 |
| $1: 00$ | 24.95 | -0.05 |
| $1: 30$ | 24.80 | -0.2 |
| $2: 00$ | 24.95 | -0.05 |
| $2: 30$ | 25.21 | 0.21 |

MGTS1000 - review section I; solutions

| $3: 00$ | 24.90 | -0.1 |
| :--- | ---: | ---: |
| $3: 30$ | 24.71 | -0.29 |
| $4: 00$ | 25.31 | 0.31 |
| $4: 30$ | 25.15 | 0.15 |
| $5: 00$ | 25.20 | 0.2 |

THE TWO OCCASIONS IN WHICH THE CONTENT FELL MORE THAN 2 STD DEV FROM THE MEAN WERE AT 11.00 AND 11.30 AM, AND 3.30 AND 4.00 PM. THE PROCESS SHUTS DOWN AFTER TWO CONSECUTIVE BAGS FALL MORE THAN TWO STD DEV FROM THE MEAN AND THEREFORE THE ANSWER IS 11.30 AND 4.00.
20. The National Education Longitudinal Survey (NELS) tracks a nationally representative sample of U.S. students from eighth grade through high school and college. Research published in Chance (Winter 2001) examined the Standardized Admission Test (SAT) scores of 265 NELS students who paid a private tutor to help them improve their scores. The next table summarizes the changes in both the SAT-Mathematics and SAT-Verbal scores for these students.

|  | SAT-Math | SAT-Verbal |
| :--- | :--- | :--- |
| Mean change in score | 19 | 7 |
| Standard deviation of score changes | 65 | 49 |

a. Suppose one of the 265 students who paid a private tutor is selected at random. Give an interval that is likely to contain this student's change in the SAT-Math score.
THE INTERVAL WOULD BE $\pm 3$ STD DEV FROM THE MEAN: -176, 214
b. Repeat part a for the SAT-Verbal score. -140, 154
c. Suppose the selected student increased their score on one of the SAT tests by 140 points. Which test, the SAT-Math or SAT-Verbal, is the one most likely to have the 140-point increase? Explain. SAT MATH, IT HAS THE LARGER STD DEV.
21. Give the percentage of measurements in a data set that are above and below each of the following percentiles:
a. 75th percentile $\mathbf{2 5 \%}$ ABOVE, 75\% BELOW
b. 50th percentile 50\% ABOVE, 50\% BELOW
c. 20th percentile $\mathbf{8 0 \%}$ ABOVE, $\mathbf{2 0 \%}$ BELOW
d. 84th percentile 16\% ABOVE, 84\% BELOW

A PERCENTILE IS A VALUE THAT DIVIDES THE DATA INTO TWO PARTS. THE LOWER PART CONTAINS AT LEAST P\% AND THE UPPER PART (100-P\%) OF THE DATA.
22. Compare the $z$-scores to decide which $x$ values lie the greatest distance above the greatest distance below the mean.
a. $x=10, \mu=5, \sigma=3 ; \mathbf{Z}=(10-5) / 3=1.667$
b. $x=0, \mu=200, \sigma=100 ; \mathbf{Z}=(0-200) / 100=\mathbf{- 2}$
c. $x=ı, \mu=4, \sigma=1 ; \mathbf{Z}=(1-4) / 1=-3$ GREATEST DISTANCE BELOW
d. $x=100, \mu=50, \sigma=25 ; \mathbf{Z}=(100-50) / 25=2$ GREATEST DISTANCE ABOVE

MGTS1000 - review section I; solutions
23. Suppose that 40 and 90 are two elements of a population data set and that their $z$-scores are -2 and 3 , respectively. Using only this information, is it possible to determine the population's mean and standard deviation? If so, find them. If not, explain why it's not possible.
$\begin{aligned} & \text { WE KNOW THAT } \\ & \text { POPULATION }\end{aligned} \quad Z=\frac{x-\mu}{\sigma}$
$-2=\frac{40-\mu}{\sigma}$ AND $3=\frac{90-\mu}{\sigma}$

## AND THAT THE TWO VALUES ARE FROM THE SAME (I.E., THEY SHARE THE MEAN AND STD DEV).

WE CAN NOW SOLVE THESE TWO EQUATIONS AND
FIND THAT THE MEAN IS 60 AND THE STD DEV IS 10.
24. The U.S. Environmental Protection Agency (EPA) sets a limit on the amount of lead permitted in drinking water. The EPA Action Level for lead is .015 milligrams per liter ( $\mathrm{mg} / \mathrm{L}$ ) of water. Under EPA guidelines, if $90 \%$ of a water system's study samples have a lead concentration less than $.015 \mathrm{mg} / \mathrm{L}$, the water is considered safe for drinking. I (coauthor Sincich) received a recent report on a study of lead levels in the drinking water of homes in my subdivision. The 90th percentile of the study sample had a lead concentration of $.00372 \mathrm{mg} / \mathrm{L}$. Are water customers in my subdivision at risk of drinking water with unhealthy lead levels? Explain.
NO THEY ARE NOT AT RISK. 90\% OF THE VALUES HAVE LEAD CONCENTRATION BELOW 0.00372 AND THEREFORE THE DRINKING WATER ARE CONSIDERED SAFE.
25. In a study of how external clues influence performance, professors at the University of Alberta and Pennsylvania State University gave two different forms of a midterm examination to a large group of introductory students. The questions on the exam were identical and in the same order, but one exam was printed on blue paper and the other on red paper (Teaching Psychology, May 1998). Grading only the difficult questions on the exam, the researchers found that scores on the blue exam had a distribution with a mean of $53 \%$ and a standard deviation of $15 \%$, while scores on the red exam had a distribution with a mean of $39 \%$ and a standard deviation of 12\%. (Assume that both distributions are approximately mound shaped and symmetric.) Suppose a student is selected at random from the group of students who participated in the study and the student's score on the difficult questions is $20 \%$. Which exam form is the student more likely to have taken, the blue or the red exam? Explain. THE RED, BASED ON THE Z-SCORES OF $20 \%$ IN BOTH TESTS. FOR THE BLUE TEST Z=-2.2, FOR THE RED TEST Z=1.583. THE SMALLER Z IS THE MORE CHANCE THERE IS OF GETTING THIS VALUE IN A RANDOM SAMPLING.

[^0]MGTS1000 - review section I; solutions

What assumption, if any, did you make about the distribution of the GPAs at the university? NORMAL DISTRIBUTION
27. In some locations, radiation levels in homes are measured at well above normal background levels in the environment. As a result, many architects and builders are making design changes to ensure adequate air exchange so that radiation will not be "trapped" in homes. In one such location, 50 homes levels were measured, and the mean level was 10 parts per billion (ppb), the median was 8 ppb , and the standard deviation was 3 ppb . Background levels in this location are at about 4 ppb .
a. Based on these results, is the distribution of the 50 homes' radiation levels symmetric, skewed to the left, or skewed to the right? Why? RIGHT SKEWED; MEAN>MEDIAN
d. Suppose another home is measured at a location 10 miles from the one sampled, and has a level of 20 ppb . What is the $z$-score for this measurement relative to the 50 homes sampled in the other location? (20-10)/3=3.333
Is it likely that this new measurement comes from the same distribution of radiation levels as the other 50? Why? NO, OBSERVATIONS SHOULD HAVE Z-SCORES BETWEEN 3 AND -3. How would you go about confirming your conclusion? CREATE A DISTRIBUTION FOR THE NEW LOCATION AND COMPARE TO THE OLD ONE.
28. The U.S. Federal Trade Commission has recently begun assessing fines and other penalties against weight-loss clinics that make unsupported or misleading claims about the effectiveness of their programs. Brochures from two weight-loss clinics both advertise "statistical evidence" about the effectiveness of their programs. Clinic A claims that the mean weight loss during the first month is 15 pounds; Clinic B claims a median weight loss of 10 pounds.
a. Assuming the statistics are accurately calculated, which clinic would you recommend if you had no other information? Why? THE SECOND ONE SINCE THE MEDIAN IS LESS SENSITIVE TO EXTREME VALUES.
29. The Age Discrimination in Employment Act mandates that workers 40 years of age or older be treated without regard to age in all phases of employment (hiring, promotions, firing, etc.). Age discrimination cases are of two types: disparate treatment and disparate impact. In the former, the issue is whether workers have been intentionally discriminated against. In the latter, the issue is whether employment practices adversely affect the protected class (i.e., workers 40 and over) even though no such effect was intended by the employer (Zabell, 1989). A small computer manufacturer laid off 10 of its 20 software engineers. The ages of all engineers at the time of the layoff are below. Analyze the data to determine whether the company may be vulnerable to a disparate impact claim.

LAYOFF

| Not laid off: | 34 | 55 | 42 | 38 | 42 | 32 | 40 | 40 | 46 | 29 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Laid off: | 52 | 35 | 40 | 41 | 40 | 39 | 40 | 64 | 47 | 44 |

WE'LL CHECK THE MEDIA SINCE THERE IS ONE EXTREME OBSERVATION (X=64) IN THE LAID-OFF SET. NOT LAID-OFF MEDIAN = 40; LAID-OFF MEDIAN=40.5. THE COMPANY IS NOT VULNERABLE.


[^0]:    26. At one university, the students are given z-scores at the end of each semester rather than the traditional GPAs. The mean and standard deviation of all students' cumulative GPAs, on which the z-scores are based, are 2.7 and .5 , respectively.
    a. Translate each of the following $z$-scores to corresponding GPA:
    $z=-2.5 \rightarrow$ GPA $=1.45$
    $z=2.0 \rightarrow$ GPA $=3.7$
    $z=-1.0 \rightarrow$ GPA=2.2
    $z=.5 \rightarrow$ GPA=2.95
    b. Students with $z$-scores below -1.6 are put on probation. What is the corresponding probationary GPA? 1.9
    c. The president of the university wishes to graduate the top $16 \%$ of the students with cum laude honors and the top $2.5 \%$ with summa cum laude honors. Where (approximately) should the limits be set in terms of $z$-scores? THE TOP 16\% ARE THOSE WITH 1 Z-SCORE ABOVE THE
    MEAN; THE TOP 2.5\% ARE THOSE WITH 2 Z-SCORES ABOVE THE MEAN.
    In terms of GPAs? CUM LAUDE GPA = 3.2; SUMMA CUM LAUDE GPA = 3.7
