Each number below represents the age of a U.S. president on his first inauguration.

| 57 | 61 | 57 | 57 | 58 | 57 | 61 | 54 | 68 | 51 | 49 |
|----|----|----|----|----|----|----|----|----|----|----|
| 50 | 48 | 65 | 52 | 56 | 46 | 54 | 49 | 50 | 47 | 55 |
| 54 | 42 | 51 | 56 | 55 | 51 | 54 | 51 | 60 | 62 | 43 |
| 56 | 61 | 52 | 69 | 64 | 64 | 55 | 55 | 47 | 54 | 47 |

A) Make a stem-and-leaf plot of these ages.

B) Make a histogram of the ages of U.S. presidents on their first inauguration

using 6 classes.   **Make sure each class is equal in size.**

C) What was the mean age of the presidents on their first inauguration?

D) Which age(s) occur most frequently?

E) What is the median age?

F) Are there any outliers to the data?

15.1 The Frequency Distribution

1. **Frequency:** how many times something occurs

2. **Frequency Distribution:** A set of data listed with their frequencies (usually listed with categories or class intervals)

3. Relative Frequency Distribution:
   A set of data listed with their percentages (usually listed with categories or class intervals)

4. **Class intervals:** Is the range of a class.  In a frequency table, all class intervals should all be equal.  To identify good class intervals, find the range of the data by taking the maximum value and subtracting the minimum value.  Then divide the range into equal parts.  Usually use 8 to 10 classes to give a good presentation of the data.

Ex: A television network has asked 25 viewers to evaluate a new police drama.  The possible evaluations are (E)xcellent, (A)bove average, a(V)erage, (B)elow average, (P)oor.  After the show, the 25 evaluations were as follows:

A, V, V, B, P, E, A, E, V, V, A, E, P, B, V, V, A, A, A, E, B, V, A, B, V.

Construct a frequency table and a relative frequency table for this list of evaluations.

5. **Bar Graph:** A graphic form of data that uses bars that have space between them, where the height gives the frequency of that category or class

Ex: Draw a bar graph of the frequency distribution of TV viewers' responses from #5.

Ex: The bar graph shows the number of Atlantic hurricanes over a period of years.  Use it to answer the following questions.
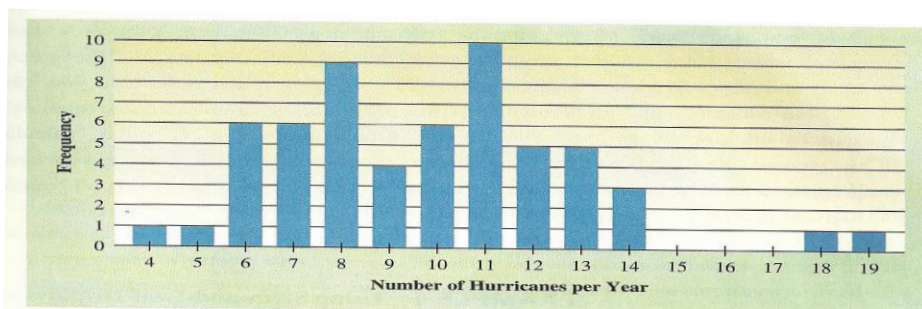


**FIGURE 14.3**   Number of hurricanes per year.

   a.   What was the smallest number of hurricanes in a year during this period?
   b.   What number of hurricanes per year occurred most frequency?
   c.   How many years were the hurricanes counted?
   d.   In what percentage of the years were there more than ten hurricanes?

6. **Back-to-back Bar Graph:**
   Plot on a two-quadrant coordinate system with the horizontal scale repeated in each direction from the central axis.

Ex: Use the side by side bar graph to answer the questions.



**Educational Attainment in 1990 versus 2006**
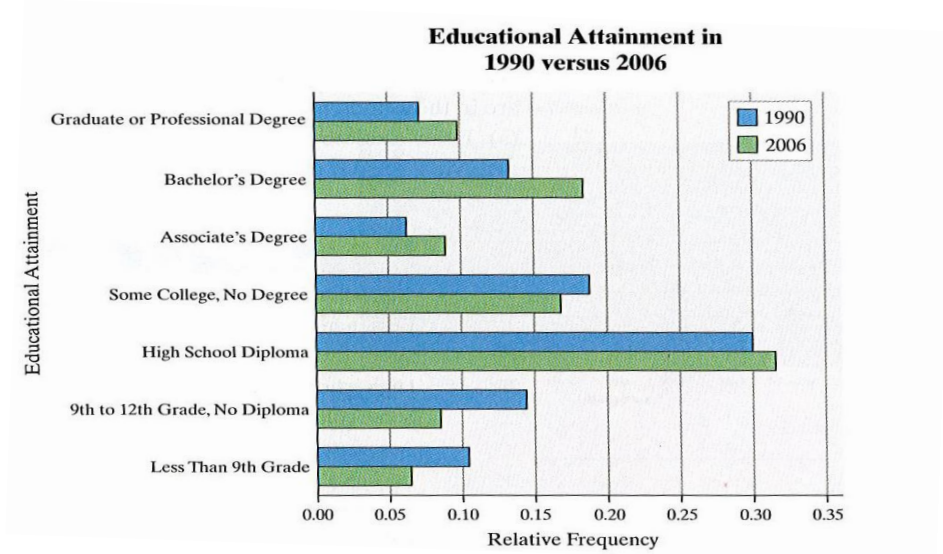
   a. What percent of Americans had their associate's degree in 1990?  2006?
   b. What percent of Americans at least had a high school diploma in 1990?  2006?
   c. Are a greater proportion of Americans dropping out of college before earning a degree?

**Another view of the same data.**



**Educational Attainment, 2006**
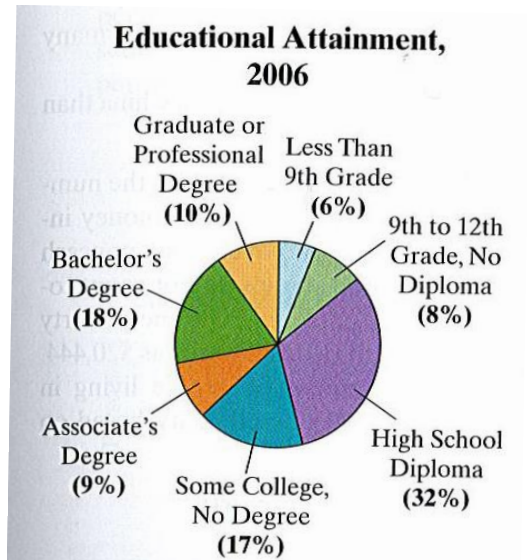
7. **Histogram:** A type of bar graph in which the width of each bar represents a class interval and the height of the bar represents the frequency in that interval

   1. Used to graph a frequency distribution of continuous variable quantity
   2. No spaces between bars of classes, usually have less than ten intervals

Ex: The New York Clinic has the following data regarding the weight lost by its clients over the past six months. Draw a histogram for the relative frequency distribution for this data.

| Pounds Lost | Frequency |
|---|---|
| 0-10 | 14 |
| 10 + to 20 | 23 |
| 20 + to 30 | 17 |
| 30 + to 40 | 8 |
| 40 + to 50 | 3 |

8. **Class Limit:** Is the upper and lower values in a class interval.

9. **Class Mark:** Is the midpoint of the classes.

10. The heights (in inches) of 43 girls trying out for basketball at Forest View High School have been tallied in the chart below.

| Class Limits | Tally | Frequency |
|---|---|---|
| 58-62 | || | |
| 62-66 | ⊬ || | |
| 66-70 | ⊬ ⊬ ⊬ | | |
| 70-74 | ⊬ ⊬ || | |
| 74-78 | ⊬ | |

a. Determine if the class intervals of 4 are appropriate for the data.

b. Draw a histogram of the data.

11. **Stem-and-leaf Plot:** A display of numerical data for which each value is separated into two numbers, a stem and a leaf. Leaves are only 1 unit long.

Ex: Make a stem-and-leaf plot to compare Babe Ruth's and Hank Aaron's Home Run records.
   Ruth: 54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22
   Aaron: 44, 30, 39, 40, 34, 45, 44, 32, 44, 39, 44, 38, 47, 34, 40

12. **Median:** The middle value of a data set. If you have an even number of data items, then you have to find the average of the two data items that make up the middle. Extreme values have very little influence on the median.

13. **Range:** The difference of the greatest and least values in a set of data.

14. **Quartile:** One of four groupings of a set of data determined by the median of the set and the medians of the sets determined by the median.

15. **Interquartile Range:** The difference between the third quartile point and the first quartile point.

16. **Box-and-whisker Plot:** A diagram that graphically displays the median, quartiles, extreme values, and outliers in a set of data. Whiskers extend to the extreme values of the data, unless there is an outlier When there is an outlier, then it is the extreme value.

17. **Outliers:** Extreme values that are more than 1.5 of the interquartile range beyond the upper or lower quartiles.  If outliers exist, each whisker is extended to the last value of the data that is not an outlier.

18. Use the following box plot of student test scores on last year's advanced algebra mid-year exam.
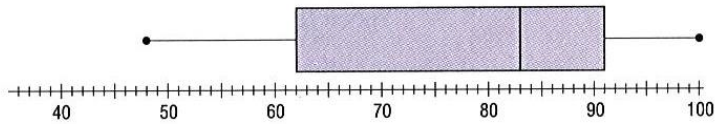


a.  What is the median score?

b.  What is the interquartile range?

c.  What percent of the students scored between 62 and 91?

d.  What is the interval of scores of students who ranked below the lower quartile?

19. The National Football League is separated into two parts-the American Football Conference (AFC) and the National Football Conference (NFC).  Here are separate box plots of the capacities of the football stadiums used by the AFC and NFC.
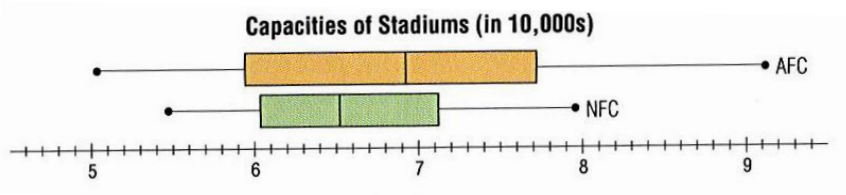


a.  What is the median capacity in each conference?

b.  What is the size of the largest stadium in each conference?

c.  About what percent of the stadiums in the AFC hold fewer than 60,000 people?

d.  On the whole, which conference has larger stadiums?

20. Below is a stem plot of the amount of money spent by 25 shoppers at a grocery store.  The stem is in $10 units.

| | |
|---|---|
| 0 | 3 8 |
| 1 | 0 1 7 8 9 |
| 2 | 0 0 3 6 8 |
| 3 | 1 3 4 7 |
| 4 | 2 5 5 |
| 5 | 0 |
| 6 | 5 |
| 7 | 2 6 |
| 8 | |
| 9 | 7 |
| 10 | |
| 11 | 3 |

a.  Find the median.

b.  Find the lower quartile.

c.  Find the upper quartile.

d.  Determine if there are any outliers.

e.  Construct a box and whiskers plot.

21. Steps in the Calculator

Homework: Bar Graphs, Histograms, and Box and Whisker Plots Worksheet

# Measures of Central Tendency:

Range: The difference of the greatest and least values and it measures variability.

Deviation: _____

Variance: _____

Standard (Mean) Deviation: _____

Population Standard Deviation versus Sample **Standard Deviation**

Population:  Divide by 'n' is used when the sample is the population

$$\sigma x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Sample:  Divide by 'n - 1' is used for a sample because it gives a better estimate of the population mean

$$Sx = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

"n" = _____

4.  Find the mean and standard deviation for the following data: {3,5,6,7,9,11,22}.  Use the table on the right.

| Value | Mean | Deviation from mean | Square of deviation from mean |
|-------|------|---------------------|-------------------------------|
|       |      |                     |                               |
|       |      |                     |                               |
|       |      |                     |                               |
|       |      |                     |                               |
|       |      |                     |                               |
|       |      |                     |                               |
|       |      |                     |                               |

**USING CALC:  Stat > Calc      1:  1-Var Stats**

 Symbol for mean: _____          Sx = sample standard deviation calculated using (n-1)

                                       σx = population standard deviation calculated using "n"

| Value | Mean | Deviation from mean | Square of Deviation from Mean |
|-------|------|---------------------|-------------------------------|
| 46    |      |                     |                               |
| 49    |      |                     |                               |
| 58    |      |                     |                               |
| 68    |      |                     |                               |
| 72    |      |                     |                               |
| 75    |      |                     |                               |
| 76    |      |                     |                               |
| 78    |      |                     |                               |
| 80    |      |                     |                               |
| 81    |      |                     |                               |
| 81    |      |                     |                               |
| 83    |      |                     |                               |
| 88    |      |                     |                               |
| 91    |      |                     |                               |
| 95    |      |                     |                               |

AFM Unit 8 Day 3 HW Worksheet                                    NAME _____

Make a box-and-whisker plot of the data.  Find the five main values needed to label your diagram.
 1. 29, 34, 35, 36, 28, 32, 31, 24, 24, 27, 34



TV Time (minutes)

0    15              60                110                      225
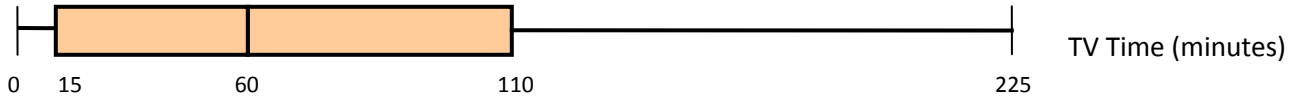
2.  What percent of the sophomores watch TV for at **least 15** minutes per night?

3.  What is the 3$^{rd}$ quartile for the TV time data?

Find the mean and standard deviation of each data set.
1.{200, 476, 721, 579, 152, 158}            2.{369, 398, 381, 392, 406, 413, 376, 454, 420, 385, 402, 446}

3.Given the quiz grades in Theresa Chair's two sections of Geometry, answer the following questions.

a.  Find the mean for 1$^{st}$ period.
b.  Find the standard deviation for 1$^{st}$ period.
   Will you use population or sample?  Why?
c.  Find the mean for 3$^{rd}$ period.
d.  Find the standard deviation for all Ms. Chair's Geometry
students. Will you use population or sample?  Why?

| 1$^{st}$ Period | | 3$^{rd}$ Period |
|---|---|---|
| | 3 | 2 9 |
| 0 | 4 | |
| 8 5 3 | 5 | |
| 9 6 5 2 0 | 6 | |
| 5 5 2 0 | 7 | 8 |
| 5 2 1 | 8 | 0 5 8 8 8 9 |
| | 9 | 0 2 4 5 5 5 6 |
| 0 0 0 | 10 | 0   3 5 8 |

Determine whether the data in each table appear to be left-skewed, right-skewed, or normally distributed, or bimodal

1.U.S. Population        5. Record Low Temperatures in the 50 States              6.  GPAs of Jr at Apex

| Age | Percent |
|---|---|
| 0-19 | 28.7 |
| 20-39 | 29.3 |
| 40-59 | 25.5 |

| Temperature (°F) | Number of States |
|---|---|
| $-80\ to-65$ | 4 |
| $-64\ to-49$ | 12 |
| $-48to-33$ | 19 |

| GPA | Frequency |
|---|---|
| 0.0-0.4 | 4 |
| 0.5-0.9 | 4 |
| 1.0-1.4 | 2 |
| 1.5-1.9 | 32 |

6.

| Shoe Size | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|

$-16\ to-1$

7.Time Spent in a Museum              8. Average Age of High School Principals    9.

0 to15

| Minutes | Frequency |
|---|---|
| 0-25 | 27 |
| 26-50 | 46 |
| 51-75 | 89 |

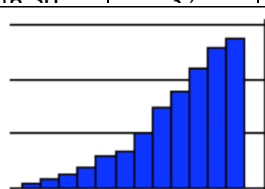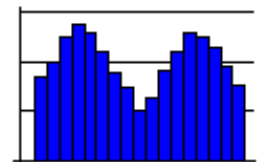| Age in Years | Number |
|---|---|
| 31-35 | 3 |
| 36-40 | 8 |
| 41-45 | 15 |
| 46-50 | 32 |
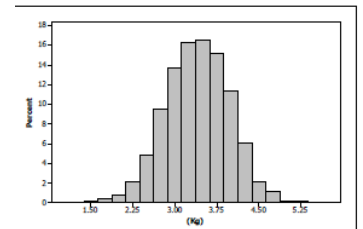


10.
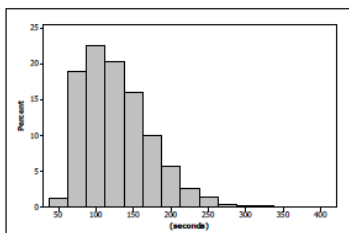


33

11.



12.

Sampling Techniques and Misuses

1. **Statistics:**
   The science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions.  In addition, statistics is about providing a measure of confidence in any conclusions.

2. **Population:**
   The entire group of individuals that we want information about

3. **Sample:**
   A part of the population that we actually examine in order to gather information.  The sample is used to make generalizations of the population.

4. **Qualitative Data:**
   Data that measures the classification of individuals based on some attribute or characteristic

5. **Quantitative Data:**
   Data that provides numerical measures of individuals.

Ex: Determine whether the following variables are qualitative or quantitative.
   a. Gender
   b. Temperature
   c. Number of days during the past week that a college student
      aged 21 years or older has had at least one drink.
   d. Zip code

6. **Discrete Variable:**
   A quantitative variable that has either a <u>finite number</u> of possible values or a countable number of values

7. **Continuous Variable:**
   A quantitative variable that has an <u>infinite number</u> of possible values that are not countable.

Ex: Determine whether the following quantitative variables are discrete or continuous.

   a. The number of heads obtained after flipping a coin five times.

   b. The number of cars that arrive at a McDonald's drive-through
      between 12:00 P.M. and 1:00 P. M.

   c. The distance a 2007 Toyota Prius can travel in city driving conditions with a full tank of gas.

8. How do we gather data?
   - Surveys
   - Observational Studies
   - Experimental Studies
   - Simulations

9. **Observational Study:**
   Investigators observe subjects and measure variables of interest without assigning treatments to the subjects. The treatment that each subject receives is determined beyond the control of the investigator. *Do not allow a researcher to claim causation, only association.

10. **Experimental Study:**
    Investigators apply treatments to experimental units (people, animals, plots of land, etc.) and then proceed to observe the effect of the treatments on the experimental units.

11. **Simulations:**
    The use of a mathematical model to recreate a situation, often repeatedly, so that the likelihood of various outcomes can be more accurately estimated.

Ex:  Identify each as an observational study, experimental study, survey or simulation.
   a. The muscles of men aged 40 - 50 were 40% to 50% stronger after they participated in a 10 week, high-intensity, resistance training program twice a week.
   b. On average, suppose a baseball player hits a home run once in every 10 times at bat, and suppose he gets exactly two "at bats" in every game.  Estimate the likelihood that the player will hit 2 home runs in a single game.
   c. Forty volunteers suffering from insomnia were divided into two groups.  The first group was assigned to a special no-desserts diet while the other continued desserts as usual.  Half of the people in these groups were randomly assigned to an exercise program, while the others did not exercise. Those who ate no desserts and engaged in exercise showed the most improvement.
   d. In 2001, a report in the *Journal of the American Cancer Institute* indicated that women who work nights have a 60% greater risk of developing breast cancer.  Researchers based these findings on the work histories of 763 women with breast cancer and 741 women without the disease.
   e. Scientists at a major pharmaceutical firm investigated the effectiveness of an herbal compound to treat the common cold.  They exposed each subject to a cold virus, and then gave him or her either the herbal compound or a sugar solution known to have no effect.  Several days later, they assessed the patient's condition, using a cold severity scale of 0 to 5.

12. **Sampling Design:** refers to the <u>method</u> used to choose the sample from the population

13. **Sampling Frame:** a list of every individual in the population

14. **Simple Random:**
   - consist of $n$ individuals from the population chosen in such a way that every individual has an equal chance of being selected

   Suppose we were to take a Simple Random Survey (SRS) of 100 Apex students – Put each students' name in a hat.  Then randomly select 100 names from the hat.  Each student has the same chance to be selected!

   - Not only does each student have the same chance to be selected – but every possible group of 100 students has the same chance to be selected!  Therefore, it has to be possible for all 100 students to be seniors in order for it to be an SRS!
   - Advantages: Unbiased
   - Disadvantages: Large variance, may not be representative. It must have a sampling frame (list of population)

**15. Stratified random sample:**
- population is divided into homogeneous groups called strata
  Homogeneous groups are groups that are alike based upon some characteristic of the group members.
- SRS's are pulled from each strata

Suppose we were to take a stratified random sample of 100 APEX students. Since students are already divided by grade level, grade level can be our strata. Then randomly select 50 seniors and randomly select 50 juniors.
- Advantages: More precise unbiased estimator than SRS, Less variability, Cost reduced if strata already exists
- Disadvantages: Difficult to do if you must divide stratum, Formulas for SD & confidence intervals are more complicated, Need sampling frame

**16. Cluster Sample:**
- based upon location
- randomly pick a location & sample all there

Suppose we want to do a cluster sample of APEX students. One way to do this would be to randomly select 10 classrooms during $2^{nd}$ period. Sample all students in those rooms!
- Advantages: Unbiased, Cost is reduced, Sampling frame may not be available (not needed)
- Disadvantages: Clusters may not be representative of population, Formulas are complicated

**17. Convenience Sample:**
- Ask people who are easy to ask
- Produces bias results

An example would be stopping friendly-looking people in the mall to survey. Another example is the surveys left on tables at restaurants - a convenient method!

The data obtained by a convenience sample will be biased – however this method is often used for surveys & results reported in newspapers and magazines!

Ex: Identify the sampling design.
  a. The Educational Testing Service (ETS) needed a sample of colleges. ETS first divided all colleges into groups of similar types (small public, small private, etc.) Then they randomly selected 3 colleges from each group.
  b. A county commissioner wants to survey people in her district to determine their opinions on a particular law up for adoption. She decides to randomly select blocks in her district and then survey all who live on those blocks.

**18. Bias:**
- A systematic error in measuring the estimate
- favors certain outcomes
  Anything that causes the data to be wrong! It might be attributed to the researchers, the respondent, or to the sampling method!

Sampling Design Worksheet                    Name: _____

**I.  Classify the variable as qualitative or quantitative.**

1.  Number of siblings
2.  Grams of carbohydrates in a doughnut
3.  Number on a football player's jersey
4.  Number of unpopped kernels in a bag of ACT microwave popcorn
5.  Assessed value of a house
6.  Phone number
7.  Student ID number

**II.  Determine whether the quantitative variable is discrete or continuous.**

8.  Runs scored in a season by Albert Pujols
9.  Volume of water lost each day through a leaky faucet.
10. Length (in minutes) of a country song
11. Number of sequoia trees in a randomly selected acre of Yosemite National Park
12. Temperature on a randomly selected day in Memphis, Tennessee
13. Internet connection speed in kilobytes per second
14. Points scored in an NCAA basketball game
15. Air pressure in pounds per square inch in an automobile tire

**III. Determine whether the study depicts an observation study, experimental study, simulation or a survey.**

16. Researchers wanted to know if there is a link between proximity to high-tension wires and the rate of leukemia in children.  To conduct the study, researchers compared the incidence rate of leukemia for children who lived within ½ mile of high-tension wires to the incidence rate of leukemia for children who did not live within ½ mile of high-tension wires.

17. Rats with cancer are divided into two groups.  One group receives 5 milligrams (mg) of a medication that is thought to fight cancer, and the other receives 10 mg.  After 2 years, the spread of the cancer is measured.

18. Rolling a pair of dice to determine the chance of getting a five or a two.

19. Seventh-grade students are randomly divided into two groups.  One group is taught math using traditional techniques; the other is taught math using a reform method.  After 1 year, each group is given an achievement test to compare proficiency.

20. Using a coin to estimate the percent of families that have one boy when there is three children.

21. A poll is conducted in which 500 people are asked whom they plan to vote for in the upcoming election.


22. A survey is conducted asking 400 people, "Do you prefer Coke or Pepsi?"


23. While shopping, 200 people are asked to perform a taste test in which they drink form two randomly placed, unmarked cups. They are then asked which drink they prefer.


24. Conservation agents netted 250 large-mouth bass in a lake and determined how many were carrying parasites.


**IV. Determine the sample design.**

25. To determine customer opinion of its boarding policy, Southwest Airlines randomly selects 60 flights during a certain week and surveys all passengers on the flights.


26. In an effort to identify if an advertising campaign has been effective, a marketing firm conducts a nationwide poll by randomly selecting individuals from a list of known users of the product.


27. A farmer divides his orchard into 50 subsections, randomly selects 4, and samples all the trees within the 4 subsections to approximate the yield of his orchard.


28. To determine his DSL Internet connection speed, Shawn divides up the day into four parts: morning midday, evening, and late night. He then measures his Internet connection speed at 5 randomly selected times during each part of the day.


29. A large medical professional organization with membership consisting of doctors, nurses, and other medical employees wanted to know how its members felt about HMOs (health maintenance organizations). Name the type of sampling plan they used in each of the following scenarios:

    a. They randomly selected 500 members from each of the lists of all doctors, all nurses, and all other employees and surveyed those 1500 members.


    b. They randomly selected ten cities from all cities in which its members lived, and then surveyed all members is those cities.


30. A political scientist surveys 400 voters randomly selected from the list of all registered voters in a community. The purpose is to estimate the proportion of registered voters who will vote in an upcoming election.

15.4   The Normal Distribution

**1. Normal Distribution:** A frequency distribution that often occurs when there is a large number of values in a set. The distribution of data along a bell-shaped, symmetric curve that reaches its maximum height at the mean.

**2. Properties of Normal Curve:**

> **68.3 % of the distribution is within 1 standard deviation of the mean.**
> **95.5 of the distribution is within 2 standard deviations of the mean.**
> **99.7 % of the distribution is within 3 standard deviations of the mean.**

**3. Standard Normal Distribution:** The normal distribution having a mean of 0 and a standard deviation of 1. It is particularly important because any normal distribution can be related to it through the use of standard values. The total area under the curve (and above the x-axis) is 1. Also, the area under the curve to the left of the number z is the proportion of the data having standard values less than z.

Ex: Suppose the scores of 500 college freshmen taking Psychology 101 are normally distributed. The mean score is 60 out of 100, and the standard deviation is 10. Sketch a normal curve that represents the frequency scores.

Using the example, estimate how many grades will fall between each of the following intervals:

a.   50 – 70                    b. 40 – 80                    c. 30 – 90

d.   55 – 65                    e. 45 – 75

Using the above example, estimate each probability:

    a. P(grade < 72)                           b. P(grade > 85)                   c. P(73 < grade < 83)

Ex: Find the upper and lower limits of the interval about the mean in which 72% of the values of a set of normally distributed data can be found if the mean is 124 and the standard deviation is 16.

Ex: In a normal distribution, about what percent of the data:

    a.  fall within 2.2 standard deviations of the mean?

    b.  do not fall within 2.2 standard deviations of the mean?

Ex: A day is selected at random at a post office whose daily letter-handling rate is normally distributed. The mean number of letters per day is 10,000 and the standard deviation is 350. What is the probability that the post office hand les between 9000 and 11,000 letters per day?

1. In a certain large school district, the set of all standardized mathematics scores is normally distributed with mean $\bar{x} = 540$ and standard deviation of 64. What is the probability that a student chosen at random scores between 580 and 620 on that test?

2. Suppose 300 values in a set of data were normally distributed.

   a. How many values are within one standard deviation of the mean?

   b. How many values are two standard deviations of the mean?

   c. How many values fall in the interval between the mean and one standard deviation above the mean?

3. **Percentiles:** The value of a variable below which a certain <u>percent</u> of observations fall.

   For example, the 20th percentile is the value (or score) below which 20 percent of the observations may be found.

   The term percentile and the related term <u>percentile rank</u> are often used in the reporting of scores from <u>norm-referenced tests</u>.

   For example, if a score is in the 86th percentile, it is higher than 85% of the other scores.

Normal Distribution Worksheet

1.Mrs. Sung gave a test in her trigonometry class.  The scores were normally distributed with

   mean of 85 and a standard deviation of 3.

     a.   What percent would you expect to score between 82 and 88?

     b.   What percent would you expect to score between 88 and 91?

     c.   What is the probability that a student chosen at random scored between 79 and 91?

2. The useful life of a radial tire is normally distributed with a mean of 30,000 miles and a standard deviation of

5000 miles.  The company makes 10,000 tires a month.

     a.   About how many tires will last between 25,000 and 35,000 miles?

     b.   About how many tires will last more than 40,000 miles?

     c.   About how many tires will last less than 25,000 miles?

     d.   What is the probability that if you buy a radial tire at random, it will last between 20,000 and 35,000
        miles?

3. The shelf life of a particular dairy product is normally distributed with a mean of 12 days and standard deviation
of 3.0 days.

     a.   About what percent of the products last between 9 and 15 days?

     b.   About what percent of the products last between 12 and 15 days?

     c.   About what percent of the products last less than 3 days?

     d.   About what percent of the products last more than 15 days?

4. Suppose that for a particular brand of television set, the distribution of picture tube failures has a mean of 120
months with a standard deviation of 12 months.  What percent of tubes will last less than 140 months?  Or more
than 140 months?

5. Assume that in the NBA the distribution of heights has a mean of 6 feet, 8 inches, and a standard deviation of 3
inches.  If there are 324 players in the league, how many players do you expect to be over seven feet tall?  (watch
units here!)

6. Assume that the distribution of 21-year-old women's heart rates at rest is 68 beats per minute with a standard
deviation of 4 beats per minute.  If 200 women are examined, how many would you expect to have a heart rate of
less than 70?

# Normal Distribution Worksheet

1. A set of 1000 values has a normal distribution.  The mean of the data is 120, and the standard deviation is 20.
   a. How many values are in one standard deviation from the mean?
   b. What percent of data is in the range 110 to 130?
   c. What percent of the data is in the range 90 to 110?
   d. Find the range about the mean that includes 90% of the data?
   e. Find the range about the mean that includes 70% of the data?
   f. Find the probability that a value selected at random from the data will be within the limits 100 and 150?
   g. Find the probability that a value selected at random from the data will be greater than 140?
   h. Find the point below which 90% of the data lie?

2. X is a normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find

a) $P(x < 40)$          b) $P(x > 21)$          c) $P(30 < x < 35)$

3. A radar unit is used to measure speeds of cars on a motorway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr?

4. For a certain type of computers, the length of time between charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours.

5. Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university?

6. The length of similar components produced by a company are approximated by a normal distribution model with a mean of 5 cm and a standard deviation of 0.02 cm. If a component is chosen at random:
 a) What is the probability that the length of this component is between 4.98 and 5.02 cm?

 b) What is the probability that the length of this component is between 4.96 and 5.04 cm?

7. The length of life of an instrument produced by a machine has a normal distribution with a mean of 12 months and standard deviation of 2 months. Find the probability that an instrument produced by this machine will last:
 a) less than 7 months.          b) between 7 and 12 months.

8. The time taken to assemble a car in a certain plant is a random variable having a normal distribution with a mean of 20 hours and a standard deviation of 2 hours. What is the probability that a car can be assembled at this plant in a period of time:
 a) less than 19.5 hours?          b) between 20 and 22 hours?

9. A large group of students took a test in Physics and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the students:
 a) scored higher than 80?      b) should pass the test (grades≥60)?      c) should fail the test (grades<60)?

10. The annual salaries of employees in a large company are approximately normally distributed with a mean of $50,000 and a standard deviation of $20,000.

a) What percent of people earn less than $40,000?

b) What percent of people earn between $45,000 and $65,000?

c) What percent of people earn more than $70,000?

11. In a city, it is estimated that the maximum temperature in June is normally distributed with a mean of 23° and a standard deviation of 5°. Calculate the number of days in this month in which it is expected to reach a maximum of between 21° and 27°.

12. The mean weight of 500 college students is 70 kg and the standard deviation is 3 kg. Assuming that the weight is normally distributed, determine how many students weigh:
 a) Between 60 kg and 75 kg.          b) More than 80 kg.          c) Less than 64 kg.

**NORMAL DISTRIBUTION- WORKSHEET #2**

1. Twelve students were given a AFM quiz and the times (in minutes) they took to complete the quiz are shown: 8,12,15,16,14,10,10,11,13,15,9,11. Find the range, mean , mode, median, Q1, Q3, IQR and the standard deviation.

For a value to be an outlier it must be greater than _____ but less than _____.

2. Eight students were asked how many hours it took them to write a research paper. The times (in hours) are shown: 6, 10, 3, 5, 7, 8, 2, 7.Find the range, mean , and standard deviation.

3. To qualify to attend a fire academy, an applicant must take a written exam. If the mean of all the test scores is 80 and the standard deviation is 5, find the probability that a randomly selected applicant scores between 75 and 95. Assume that the test scores are normally distributed.

4. The average time it takes an emergency service to respond to calls in certain municipality is 13 minutes. If the standard deviation is 3 minutes, find the probability that for a randomly selected call, the service takes less than 10 minutes. Assume the times are normally distributed.

5. If the measure of systolic blood pressure is normally distributed with a mean of 120 and a standard deviation of 10, find the probability that a randomly selected person will have a systolic blood pressure below 140. Assume systolic blood pressure is normally distributed.

6. If an automobile gets an average of 25 miles per gallon on a trip and the standard deviation is 2 miles per gallon, find the probability that on randomly selected trip, the automobile will get between 21 and 29 miles per gallon. Assume that the variable is normally distributed.

7. If adult Americans spent on average of $60 per year for books and the standard deviation of the variable is $5, find the probability that a randomly selected adult spent between $50 and $65 last year on books. Assume the variable is normally distributed.

Normal Distribution Worksheet Answers

1. a. 39.07%        b. 48.03%        c. 15.93%        d. 8.0%       e. 12.84%    f. 4.4%
          g. 79.95%    h. 11.12%

2. a. 68.3%        b. 13.61%        c. 95.46%

3. a. 6830        b. 228        c. 1587        d. 81.85%

4. a. 68.3%        b. 34.15%        c. 0.13%        d. 15.87%

5. less than 95.25%; more than 4.75%

6. 29.7 = 30    Hint: convert mean into inches! So that the mean and standard deviation are in the same units.

7. 138.3 = 138

Statistics Test Review Sheet

**I.  The table shows the resulting scores of the final exam in AFM for 90 Students at Apex High School.**

| Scores | 53-61 | 61-69 | 69-77 | 77-85 | 85-93 | 93-101 |
|--------|-------|-------|-------|-------|-------|--------|
| Frequency | 3 | 7 | 11 | 38 | 19 | 12 |

1. Find the class limits for the row with frequency 19
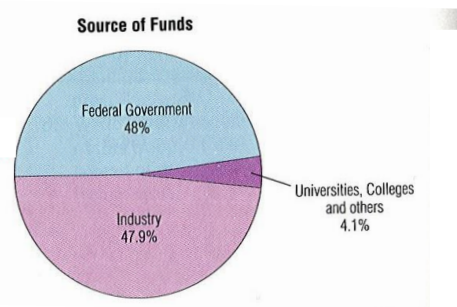2. Find the class interval
3. Make a histogram of the data

**II.  The data shows the average temperatures In January for several cities in the mid- South**
   **49.1  50.8  42.9  44  44.2  51.4  45.7  39.9  50.8  46.7  52.4  50.4**
4. Find the mean, median and the mode
5. Find the values of Q1,Q2,Q3,range,  interquartile range and  the outliers.
6.  Sketch a Box-and –Whisker plot.

**III. A family goes grocery shopping every week. In a month the cost of the groceries are :**
   **$72.42   $91.5   $58.99   $69.02**
7. Find the mean.
8. Find the standard deviation. (which are you going to use: population or sample?)

**IV. Refer to the circle graph which shows who paid for the research and development done in the U. S. in 1988.  If the total amount of research funded was about $126,115 million:**



Source of Funds
Federal Government 48%
Industry 47.9%
Universities, Colleges and others 4.1%

9.  About how much was funded by the Federal government?
10. About how much was funded by industry?

**V.  Use the histogram to answer the following:**

11. Describe the shape of the distribution.

12. Which is higher the mean or the median?

13. About what percentage of the scores are greate than or equal to 15 and less than 18?
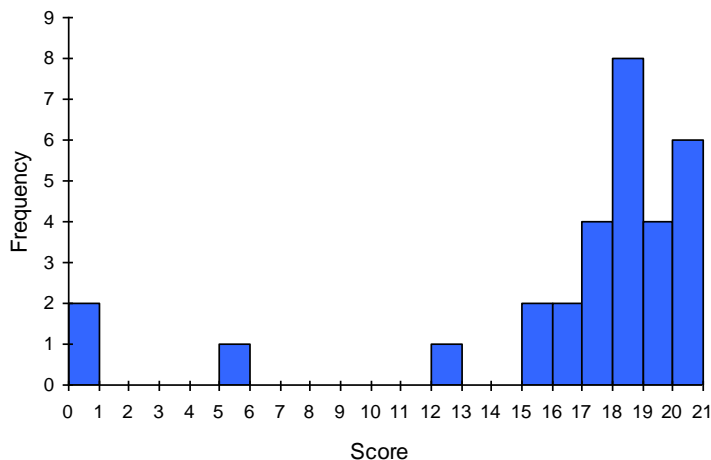
**VI. In a group of 230 tests , the scores are normally distributed with a mean grade average of 76 and a standard deviation of 10.**

14. Sketch a normal curve representing the data.  Label the points that would be 1, 2 and 3 standard deviations from the mean.
15. How many students score above 96?
16. How many students score below 66?
17. How many students score within one standard deviation of the mean?
18. What percent of the data is in the interval 71 – 101?
19. What percent of the data is in the interval 61 – 94?
20. Find the probability that the data is less than 93?
21. Find the interval about the mean that includes 86% of the data?
22. Find the limit below which 78.8 % of the data lies.
23. How many of the students would lie in the interval 69-84?
24. How many of the students would lie in the interval 86-96?

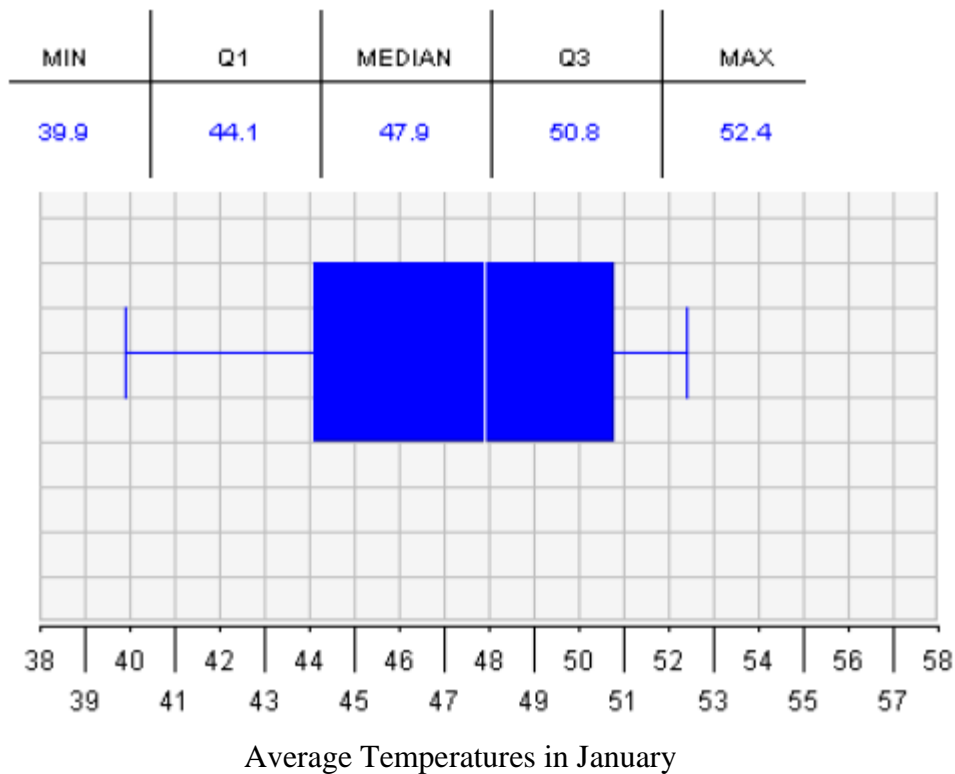VII.     More Normal Distribution practice.
25. In order to qualify for a position, an applicant must score 85 or above on a skills test. If the test scores are normally distributed with a mean of 80 and a standard deviation of 3, find the probability that a randomly selected applicant will qualify for the position.

26  If a brisk walk at 4 miles per hour burns an average of 300 calories per hour, find the probability that a person will burn between 260 and 290 calories if the person walks briskly for one hour. Assume the standard deviation is 19 and the variable is normally distributed.

27. The average count for snow per year that a city receives is 40 inches. The standard deviation is 9 inches. Find the probability that next year the city will get less than 53 inches. Assume the variable is normally distributed.

28. If the average systolic blood pressure is 120 and the standard deviation is 10, find the probability that a randomly selected person will have a blood pressure less than 108. Assume the variable is normally distributed.

29. A survey found that on average adults watch 2.5 hours of television per day. The standard deviation is 0.5 hours. If 500 adults were randomly surveyed, find the number adults that watch between 2.2 and 2.8 hours per day. Assume the variable is normally distributed.

30. An exam which is normally distributed has a mean of 200 and a standard deviation of 20.  Find the percent of adults that will score above 236.

31. The average height for adult females is 64 inches. A female has a height of 66.5 inches and is taller than 97.72% of other females.   What is the standard deviation?

Statistics Test Review Sheet Answers

1. 85 and 93
2. 8
3.



Frequency

Or

Number of
Students

AFM Scores
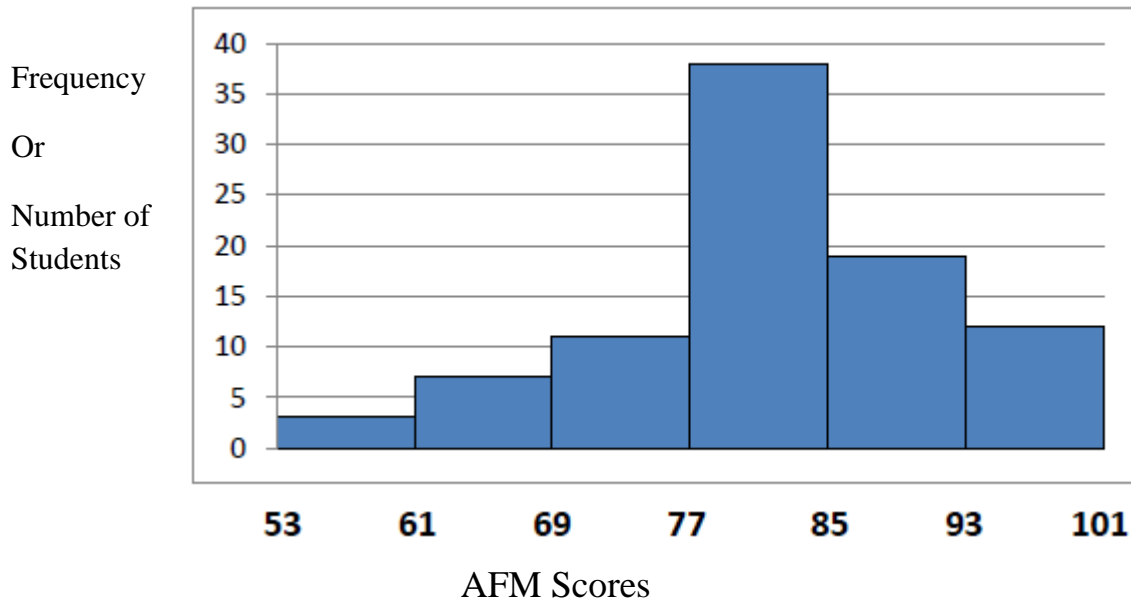
4. Median = 47.9, mode = 50.8, mean = 47.4
5. Q1 = 44.1, Q2=47.9, Q3 = 50.8, Interquartile Range = 6.7,
6.

| MIN | Q1 | MEDIAN | Q3 | MAX |
|------|------|--------|------|------|
| 39.9 | 44.1 | 47.9 | 50.8 | 52.4 |



Average Temperatures in January

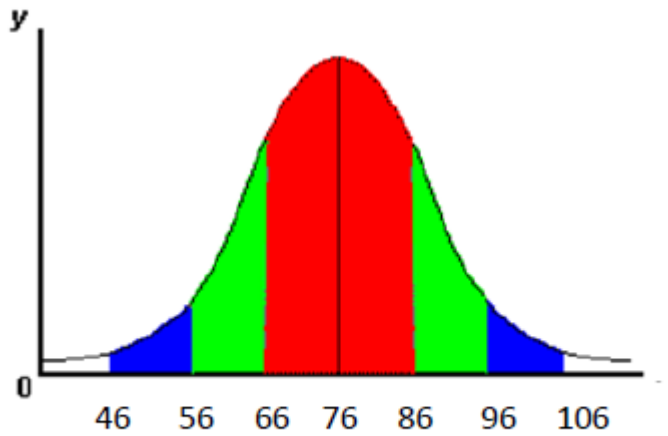7. Mean = 72.98

8. Standard deviation = 13.6
9. $60,535.20
10. $60409.09
11. Left-skewed
12. Median
13. 26.67%

14.



15. 5
16. 37
17. 157
18. 68.53%
19. 89.73%
20. 95.54%
21. 61.2-90.8
22. 84
23. 126
24. 31
25. 0.0475
26. 0.2807
27. 0.9251
28. 0.1151
29. 226
30. 3.59%
31. 1.25

AFM Statistics Review Answers

1.  a. 792.51          b. Sample          c. 1.423
2.  a. 441.8           b. Sample          c. 120.37
3.  a. 82.19           b. Population       c. 8.93
4.  a. 68.3%           b. 33.64%          c. 21.19%          d. 5.48%          e. 16.1 yrs
    f. 11.4 – 16.6     g. 8,936
5.  a. 95.5%           b. 86.98%          c. 24.51%          d. 1.68%
    e. 286.48 = 287 days                  f. 286.48 days     g. 241.12 – 290.88
6. Qualitative

7. Quantitative Discrete

8. Quantitative Continuous

9. Observational Study

10. Experimental Study

11. Simulation

12.  Convenience

13. Cluster

14. Stratified

15. Simple Random