Stats Data And Models Global 4th Edition De-Veaux Test Bank

Exam

Name_____

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Solve the problem.

1) One of the reasons that the Monitoring the Future (MTF) project was started was "to study changes in the beliefs, attitudes, and behavior of young people in the United States." Data are collected from 8th, 10th, and 12th graders each year. To get a representative nationwide sample, surveys are given to a randomly selected group of students. In Spring 2004, students were asked about alcohol, illegal drug, and cigarette use. Describe the W's, if the information is given. If the information is not given, state that it is not specified.

1) _____

• Who:
• What:
• When:
• Where:
• How:
• Why:

Answer: • Who: 8th, 10th, and 12th graders
  • What: alcohol, illegal drug, and cigarette use
  • When: Spring 2004
  • Where: United States
  • How: survey
  • Why: "to study changes in the beliefs, attitudes, and behavior of young people in the United
  States"

2) Consider the following part of a data set:

2) _____

Age (years)	Sex	Only child?	Height (inches)	Weight (pounds)	Credit Hours	GPA	Major
21	Female	Yes	67.00	140.0	16	3.60	animal science
20	Female	No	62.00	130.0	18	3.86	biology
28	Female	No	64.00	188.0	21	3.25	psychology
21	Male	No	65.00	140.0	15	2.95	psychology
24	Female	No	67.00	130.0	20	3.00	anthropology
22	Male	Yes	68.00	135.0	15	2.94	journalism

List the variables in the data set. Indicate whether each variable is treated as categorical or quantitative in this data set. If the variable is quantitative, state the units.

Answer: Categorical: sex, only child?, major
  Quantitative: age (years), height (inches), weight (pounds), credit hours, GPA

In November 2003 *Discover* published an article on the colonies of ants. They reported some basic information about many species of ants and the results of some discoveries found by myrmecologist Walter Tschinkel of the University of Florida. Information included the scientific name of the ant species, the geographic location, the depth of the nest (in feet), the number of chambers in the nest, and the number of ants in the colony. The article documented how new ant colonies begin, the ant-nest design, and how nests differ in shape, number, size of chambers, and how they are connected, depending on the species. It reported that nest designs include vertical, horizontal, or inclined tunnels for movement and transport of food and ants.

3) Describe the W's, if the information is given: 3) _____
   • Who:
   • What:
   • When:
   • Where:
   • How:
   • Why:

   Answer: • Who: Colonies of ants. "Many species of ants," but no indication of exactly how
            many.
           • What: scientific name, geographic location, average nest depth, average number of
            chambers,
             average colony size, how new ant colonies begin, the ant-nest design, and how
            nests differ
             in architecture.
           • When: November 2003
           • Where: not specified
           • How: The results of some discoveries found by myrmecologist Walter Tschinkel of
            the University
             of Florida
           • Why: Information of interest to readers of the magazine

4) List the variables. Indicate whether each variable is categorical or quantitative. If the   4) _____
   variable is quantitative, tell the units.

   Answer: Categorical: species, geographic location, how new ant colonies begin, and nest
           design.
           Quantitative: nest depth (feet), number of chambers (units), and colony size (units).

In June 2003 *Consumer Reports* published an article on some sport-utility vehicles they had tested recently. They reported some basic information about each of the vehicles and the results of some tests conducted by their staff. Among other things, the article told the brand of each vehicle, its price, and whether it had a standard or automatic transmission. They reported the vehicle's fuel economy, its acceleration (number of seconds to go from zero to 60 mph), and its braking distance to stop from 60 mph. The article also rated each vehicle's reliability as much better than average, better than average, average, worse, or much worse than average.

5) Describe the W's, if the information is given: 5) _____
   • Who:
   • What:
   • When:
   • Where:
   • How:
   • Why:

   Answer: • Who: SUV's currently on the market. We don't know how many models.
            • What: brand of vehicle, price, type of transmission, fuel economy, acceleration,
            braking distance, and reliability.
            •When: prior to June 2003
            • Where: not specified, probably the United States
            • How: testing the vehicles by driving each
            • Why: information for potential consumers

6) List the variables. Indicate whether each variable is categorical or quantitative. If the   6) _____
   variable is quantitative, tell the units.

   Answer: Categorical: brand, transmission type, reliability
            Quantitative: price (US$), fuel economy (mpg), acceleration (seconds), braking
            distance (probably
            feet?)

In July 2013, the Federal Drug Administration approved a new version of a drug used to treat opium dependence. The old version of the drug had received complaints about a bitter taste, an aftertaste, and that it took a long time to dissolve. The goal of the new version was to get more patients to take the drug as prescribed by addressing these issues. In addition to these improvements, experimenters monitored the existence and types of side effects of the drug.

7) Describe the W's, if the information is given: 7) _____
   • Who:
   • What:
   • When:
   • Where:
   • How:
   • Why:

   Answer: • Who: Patients being treated for opium addiction
            • What: Taste, aftertaste, time to dissolve, patient compliance, types of side effects
            • When: Prior to July 2013
            • Where: United States
            • How: Clinical trials
            • Why: To determine whether patients would comply with treatment better with the
            new version of the drug.

8) List the variables. Indicate whether each variable is categorical or quantitative. If the   8) _____
   variable is quantitative, tell the units.

   Answer: Categorical: taste, aftertaste, compliance, side effects
            Quantitative: time to dissolve (minutes?)

Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from "Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes" (Shanthy A. Bowman, *Journal of the American Dietetic Association*, *102*(9), pp. 1234-1239):

		Nationwide Food Survey Years			
		1987-1988	1989-1991	1994-1996	Total
Drinks Fluid Milk	Yes	354	502	366	1222
	No	226	335	366	927
	Total	580	837	732	2149

9) Find the following: 9) _____
   a. What percent of the young girls reported that they drink milk?
   b. What percent of the young girls were in the 1989-1991 survey?
   c. What percent of the young girls who reported that they drink milk were in the 1989-1991 survey?
   d. What percent of the young girls in 1989-1991 reported they drink milk?

   Answer: a. 56.9%
            b. 38.9%
            c. 41.1%
            d. 60.0%

10) What is the marginal distribution of milk consumption?   10) _____

   Answer: Yes: 1,222; No: 927

11) Do you think that milk consumption by young girls is independent of the nationwide   11) _____
   survey year? Use statistics to justify your reasoning.

   Answer: No. 56.9% of all young girls surveyed reported drinking milk, but 60% of the young girls reported drinking milk in the 1989-1991 survey. Since these percentages differ, milk consumption and year are not independent.

12) Consider the following pie charts of the a subset of the data above:

Pie Chart of 1989-1991, 1994-1996 vs Milk?

1989-1991*Milk?   1994-1996*Milk?

| Category |
| Yes |
| No |
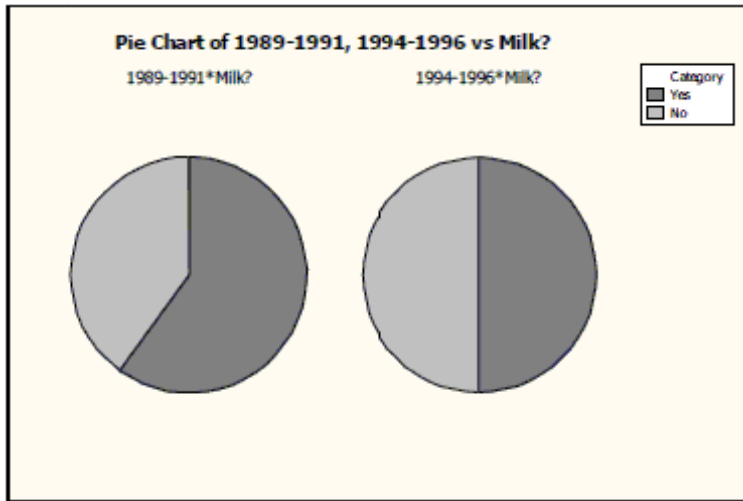
Do the pie charts above indicate that milk consumption by young girls is independent of the nationwide survey year? Explain.

Answer: No. It looks like there is some sort of relationship between milk consumption and nationwide survey year, since the percentage of young girls who reported drinking milk is a larger slice of the pie chart for the 1989-1991 survey than the same response for the 1994-1996 survey.

To determine if people's preference in dogs had changed in the recent years, organizers of a local dog show asked people who attended the show to indicate which breed was their favorite. This information was compiled by dog breed and gender of the people who responded. The table summarizes the responses.

	Female	Male	Total
Yorkshire Terrier	73	59	132
Dachshund	49	47	96
Golden Retriever	58	33	91
Labrador	37	41	78
Dalmatian	45	28	73
Other breeds	86	67	153
Total	348	275	623

13) Identify the variables and tell whether each is categorical or quantitative.

Answer: Gender and Breed; both categorical.

14) Which of the W's are unknown for these data?

Answer: We do not know how or when the people were surveyed, or where the local dog show was located.

15) Find each percent.
    a. What percent of the responses were from males who favor Labradors?
    b. What percent of the male responses favor Labradors?
    c. What percent of the people who choose Labradors were males?

    Answer: a. 6.6%
              b. 14.9%
              c. 52.6%

15) _____

16) What is the marginal distribution of breeds?

    Answer: There were 132 Yorkshire terrier responses, 96 Dachshund responses, 91 Golden Retriever
    responses, 78 Labrador responses, 73 Dalmatian responses, and 153 Other responses.

16) _____

17) Write a sentence or two about the conditional relative frequency distribution of the breeds among female respondents.

    Answer: Among females, 20.9% chose Yorkshire Terriers, 14.1% Dachshunds, 16.7% Golden Retrievers,
    10.6% Labs, and 12.9% Dalmatians. The remaining 24.7% of females preferred other breeds.

17) _____

18) Do you think the breed selection is independent of gender? Give statistical evidence to support your conclusion.

    Answer: The breed selection does not appear to be independent of gender. Overall, 56% of the respondents were females, but females were over-represented among those who favored Golden Retrievers (64%) and Dalmatians (62%), yet a much lower percentage (47%) among those who chose Labradors.

18) _____

In order to plan transportation and parking needs at a private high school, administrators asked students how they get to school. Some rode a school bus, some rode in with parents or friends, and others used "personal" transportation - bikes, skateboards, or just walked. The table summarizes the responses from boys and girls.

	Male	Female	Total
Bus	30	34	64
Ride	37	45	82
Personal	19	23	42
Total	86	102	188

19) Identify the variables and tell whether each is categorical or quantitative.
    Answer: Gender and mode of transportation, both categorical.

19) _____

20) Which of the W's are unknown for these data?
    Answer: We don't know how or when the students were surveyed, nor where the school is.

20) _____

21) Find each percent.

    a) What percent of the students are girls who ride the bus?

    b) What percent of the girls ride the bus?

    c) What percent of the bus riders are girls?

    Answer: a. 18.1%

              b. 33.3%

              c. 53.1%

21) _____

22) What is the marginal distribution of gender?

    Answer: There are 86 males and 102 females.

22) _____

23) Write a sentence or two about the conditional relative frequency distribution of modes of transportation for the boys.

    Answer: More boys (43%) caught rides to school than any other means of transportation. 35% rode the bus while only 22% used personal transportation like biking, skateboarding, or walking.

23) _____

24) Do you think mode of transportation is independent of gender? Give statistical evidence to support your conclusion.

    Answer: The way students get to school does seem to be independent of gender. Overall, 34% of students ride the bus, compared to 35% of the boys and 33% of the girls. 44% of all students caught rides with someone and 22% used personal transportation, almost the same as the percentages for boys (43% and 22%) or girls (44% and 23%) separately. These data provide little indication of a difference in mode of transportation between boys and girls at this school.

24) _____

A research company frequently monitors trends in the use of social media by American Adults. The results of one survey of 1846 randomly selected adults looked at social media use versus age group. The table summarizes the survey results.

		Age Group				
		18-29	30-49	50-64	65+	Total
Uses Social Media	Yes	328	417	288	114	1147
	No	67	125	265	242	699
	Total	395	542	553	356	1846

25) Identify the variables and tell whether each is categorical or quantitative.

    Answer: Age is numerical, but the grouping treats it as categorical, and social media use is categorical.

25) _____

26) Which of the W's are unknown for these data?

    Answer: We don't know when the adults were surveyed.

26) _____

27) Find each percent.

    a) What percent of adults surveyed are social media users aged 30-49?

    b) What percent of the social media users are aged 30-49?

    c) What percent of adults aged 30-49 are social media users?

    Answer: a. 22.6%

              b. 36.4%

              c. 76.9%

27) _____

28) What is the marginal distribution of age groups?

Answer: There were 395 adults aged 18-29, 542 aged 30-49, 553 aged 50-64, and 356 that were 65 or older.

29) Write a sentence or two about the conditional relative frequency distribution of ages of social media users.

Answer: More social media users in the survey (36.4%) were aged 30-49 than any other age group. Next was the 18-29 age group at 28.5%, then the 50-64 group at 25.1%, and the smallest group of social media users (9.6%) was the 65 and older group.

30) Do you think social media use is independent of age? Give statistical evidence to support your conclusion.

Answer: Social media use does not appear to be independent of age. Overall, 21.4% of adults surveyed were 18-29 years old, but 28.6% of social media users are in that age group. And 19.2% of adults surveyed were 65 and older, but only 9.9% of social media users were. In general, older groups seem to be underrepresented among social media users.

A survey conducted in a college intro stats class asked students about the number of credit hours they were taking that quarter. The number of credit hours for a random sample of 16 students is given in the table.

| 10 | 10 | 12 | 14 | 15 | 15 | 15 | 15 |
| 17 | 17 | 19 | 20 | 20 | 20 | 20 | 22 |

31) a. Sketch a histogram of these data
   b. Find the mean and standard deviation for the number of credit hours.
   c. Find the median and IQR for the number of credit hours.
   d. Is it more appropriate to use the mean and standard deviation or the median and IQR to summarize theses data? Explain.

Answer: a.
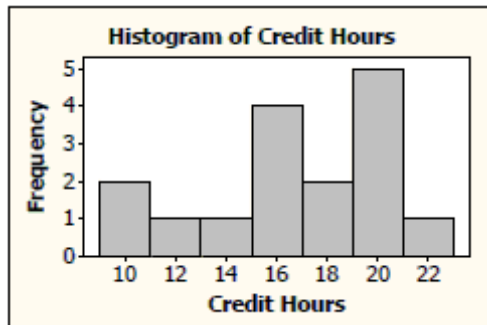
b. $\bar{x}$ = 16.3 credit hours; s = 3.7 credit hours
c. The median is 16.0 credit hours.
   IQR = Q3 - Q1 = 20 - 14.5 = 5.5 credit hours
d. It is more appropriate to use the median and IQR to summarize these data, because these data are not unimodal and symmetric.

32) Suppose that the student taking 22 credit hours in the data set was actually taking 28 credit hours instead of 22 (so we would replace the 22 in the data set with 28). Indicate whether changing the number of credit hours for that student would make each of the following summary statistics increase, decrease, or stay about the same:
a. mean
b. median
c. range
d. IQR
e. standard deviation

Answer: a. increase
             b. stay about the same
             c. increase
   d. stay about the same
             e. increase

Solve the problem.

33) The students in a biology class kept a record of the height (in centimeters) of plants for a class experiment.

49	67	38	55	62
54	36	41	56	43
48	75	44	60	48
52	48	53	59	32

a. Sketch a histogram for these data.
b. Find the mean and standard deviation of the plant heights.
c. Is it appropriate to use the mean and standard deviation to summarize these data? Explain.
d. Describe the association of plant heights.

Answer: a.
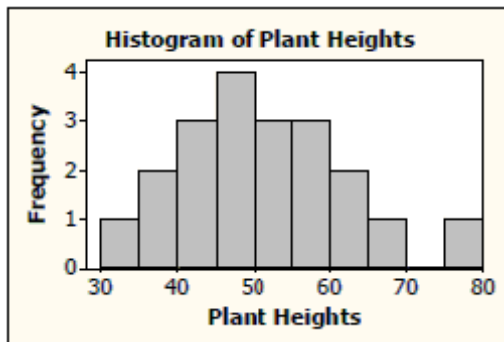
b. $\bar{x} = 51.0$ cm; $s = 10.6$ cm
c. Yes, the data are roughly unimodal and symmetric with no outliers.
d. The data are roughly symmetric with no outliers; however there is a small gap from 70 to 75 cm. The average plant height is 51.0 centimeters, with a standard deviation of 10.6 centimeters. The range of plant heights is 43 centimeters. The distribution of plant heights has a mode between 45 and 49 centimeters.

34) All students in a physical education class completed a basketball free-throw shooting event and the highest number of shots made was 32. The next day a student who had just transferred into the school completed the event, making 35 shots. Indicate whether adding the new student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:

a. mean
b. median
c. range
d. IQR
e. standard deviation

Answer: a. increase
         b. stay about the same
         c. increase
         d. stay about the same
         e. increase

35) A brake and muffler shop reported the repair bills, in dollars, for their customers yesterday.

88	283	312	290	172
154	400	381	346	181
203	118	143	252	227
56	192	292	213	422

a. Sketch a histogram for these data.
b. Find the mean and standard deviation of the repair costs.
c. Is it appropriate to use the mean and standard deviation to summarize these data? Explain.
d. Describe the association of repair costs.

Answer: a.
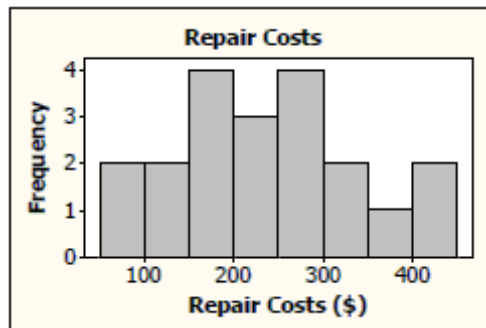
b. $\bar{x}$ = \$236.25 ; s = \$103.43
c. Yes, the data are roughly unimodal and symmetric with no outliers.
d. The repair costs averaged \$236.25, ranging from \$56 to \$422 with a standard deviation of \$103.43. The distribution was approximately symmetric, with typical repair costs clustered between \$150 and \$300.

36) On Monday, a class of students took a big test, and the highest score was 92. The next day, a student who had been absent made up the test, scoring 100. Indicate whether adding that student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:
a. mean
b. median
c. range
d. IQR
e. standard deviation

Answer: a. increase
       b. same
       c. increase
       d. same
       e. increase

There are 18 roller coasters in Virginia for which the lengths (in feet) were reported. Those lengths are listed in the table.

1385	2757	2835	2231	3828	1312
3369	3240	5100	3157	2700	600
3150	3108	2705	4210	4882	1960

37) a. Sketch a histogram for these data.
    b. Find the mean and standard deviation of the roller coaster lengths.
    c. Is it appropriate to use the mean and standard deviation to summarize these data? Explain.
    d. Describe the distribution of roller coaster lengths.

Answer: a.
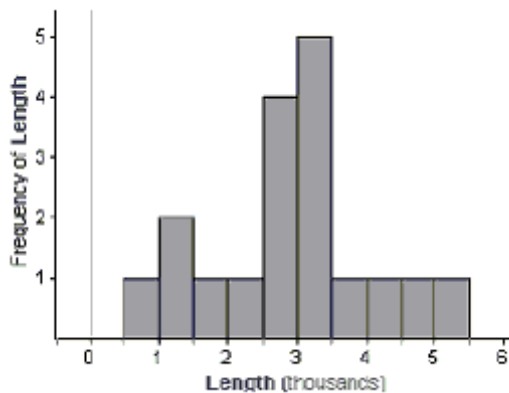
b. $\bar{x} = 2918.3$ft, $s = 1172.92$ ft
c. Yes, the data are roughly unimodal and symmetric with no outliers.
d. The mean roller coaster length is 2918 ft, ranging from 600 ft to 5100 ft. The distribution is roughly symmetric, with typical lengths clustered between 2000 ft and 4000 ft.

38) There is a proposal to replace the shortest roller coaster above with one that has a length of   38) _____
1,200 ft. Indicate whether changing that roller coaster's length would make each of these
summary statistics increase, decrease, or stay about the same.
a. mean
b. median
c. range
d. IQR
e. standard deviation

Answer: a. Increase
           b. Same
           c. Decrease
           d. Same
           e. Decrease

Solve the problem.

39) The five-number summary for midterm scores (number of points; the maximum possible   39) _____
score was 50 points) from an intro stats class is:

Min	Q1	Median	Q3	Max
16.5	32	39	43.5	48.5

a. Would you expect the mean midterm score of all students who took the midterm to be
higher or lower than the median? Explain.
b. Based on the five-number summary, are any of the midterm scores outliers? Explain.

Answer: a. The mean midterm scores of all students would probably be lower than the
median. Using the 5-number summary, it appears that the data are skewed to the
left.
b. IQR = 43.5 - 32 =11.5
Q1 - 1.5IQR = 32 - 1.5(11.5) = 14.75
Q3 + 1.5IQR = 43.5 + 1.5(11.5) = 60.75
Since both the maximum and minimum scores fall between these "fences", there are
not outliers in this data set.

40) The side-by-side boxplots show the cumulative college GPAs for sophomores, juniors, and   40) _____
seniors taking an intro stats course in Autumn 2003.

Cumulative GPA
Autumn 2003
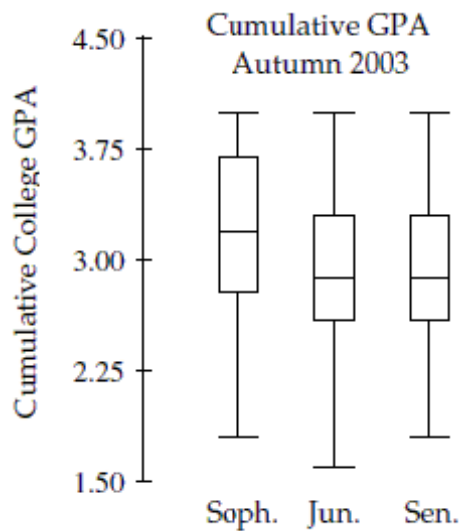
a. Which class (sophomore, junior, or senior) had the lowest cumulative college GPA?
What is the approximate value of that GPA?
b. Which class has the highest median GPA, and what is that GPA?
c. Which class has the largest range for GPA, and what is it?
d. Which class has the most symmetric set of GPAs? The most skewed set of GPAs?

Answer: a. The junior class had the lowest cumulative GPA, around 1.6.
b. The sophomore class had the highest median cumulative GPA, around 3.2.
c. The junior class had the largest range for GPA, about 2.4.
d. The senior class had the most symmetric set of GPAs. The sophomore class had
the most skewed set of GPAs, skewed to the left.

The following are histograms for the heights of 100 US women and the heights of 100 European women:
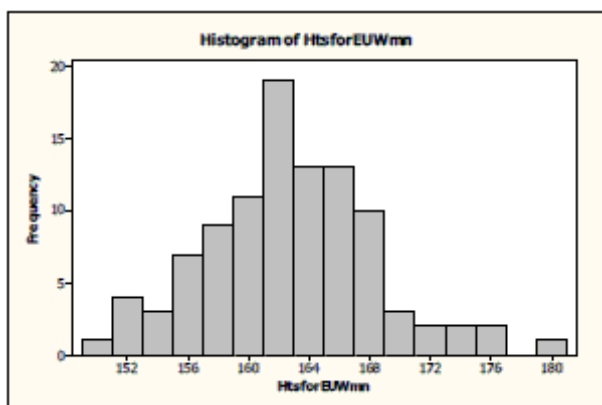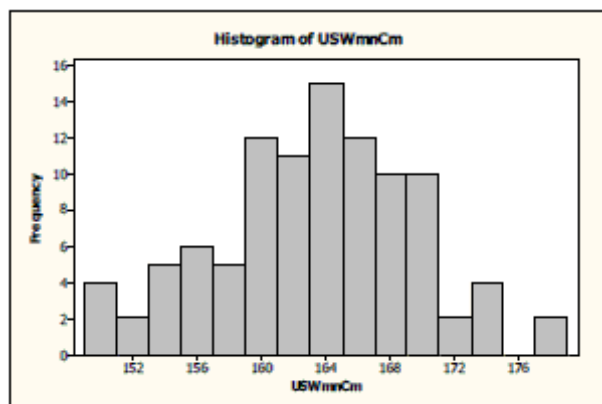
41) Compare the two distributions of the women's heights. Be sure to talk about shape, center, and spread.

41) _____

Answer: Both distributions are unimodal and roughly symmetric. Each distribution appears to be centered around 164 cm. The heights for the US women appear to be more spread out than those for the European women.

42) While the scales of the histograms are the same, there is something that could be improved so that we could compare these two distributions better. Identify this improvement and explain why it would be better.

42) _____

Answer: The frequency scales should also be the same. Additionally, the histograms could be presented either vertically or back to back, so that the horizontal scales are easily comparable.

Solve the problem.

43) The body temperature of students is taken each time a student goes to the nurse's office. 43) _____
The five-number summary for the temperatures (in degrees Fahrenheit) of students on a
particular day is:

Min	Q1	Median	Q3	Max
96.6°	97.85°	98.25°	98.6°	101.8°

a. Would you expect the mean temperature of all students who visited the nurse's office to
be higher or lower than the median? Explain.
b. After the data were picked up in the afternoon, three more students visited the nurse's
office with temperatures of 96.7°, 98.4°, and 99.2°. Were any of these students outliers
(according to the original statistics above)? Explain.

Answer: a. The mean temperature of all students would probably be higher than the median.
Using the five-number summary, it appears the data are skewed to the right.
b. IQR = 98.6° - 97.85° = 0.75°. Since 1.5(IQR) = 1.125°, the fences are 97.85° - 1.125°
= 96.725° and 98.6° + 1.125° = 99.725°. The lowest temperature (96.7°) being added to
the data set is smaller than the lower fence (96.725°) so it is an outlier on the low
end. The highest temperature (99.2°) being added to the data set is not above the
upper fence (99.725°) so it is not an outlier on the high end.

44) The boxplots show the age of people involved in accidents according to their role in the accident.
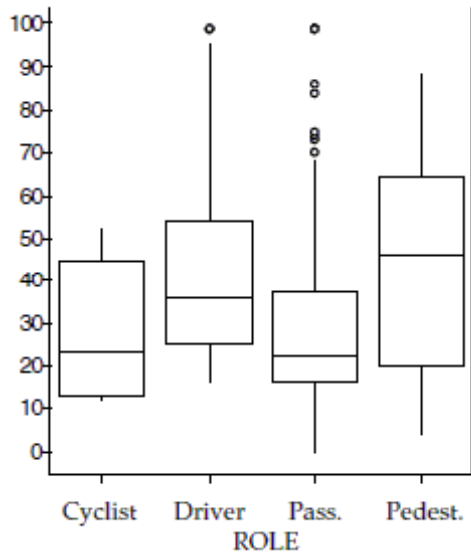
a. Which role involved the youngest person, and what is the age?
b. Which role had the lowest median age, and what is the age?
c. Which role had smallest range of ages, and what is it?
d. Which role had the largest IQR of ages, and what is it?
e. Which role generally involves the oldest people? Explain.

Answer: a. Passenger, less than 1 year.
b. Passenger, 21 yrs
c. Cyclist, 40 yrs
d. Pedestrian, 44 yrs
e. Pedestrian. While the oldest person involved in an accident is not a pedestrian, the median age for pedestrians is almost 45 years, while the median age in the other groups are between 22 and 35 years old. The oldest 50% of the Pedestrian group, from 45 to 87 years, is generally older than the youngest 75% of two groups - Cyclist and Passenger, and only the Driver group has any of its middle 50% as old. The driver and passenger groups have a few people older than the pedestrian group.

45) One thousand students from a local university were sampled to gather information such as   45) _____
gender, high school GPA, college GPA, and total SAT scores. The results were used to
create histograms displaying high school grade point averages (GPAs) for both males and
females. Compare the grade distribution of males and females.
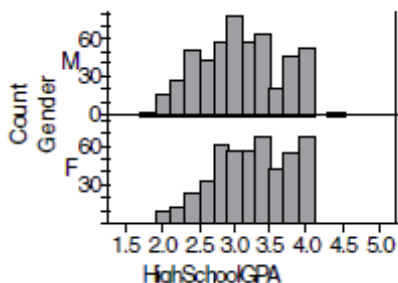
Answer: The distributions of high school GPA for both males and females are skewed to the
left, and both distributions appear to be centered at a GPA of about 3.0. The
distribution of male GPA appears slightly more spread out than the distribution of
female GPA.

46) The five-number summary for the weights (in pounds) of fish caught in a bass tournament   46) _____
is:

Min	Q1	Median	Q3	Max
2.3	2.8	3.0	3.3	4.5

a. Would you expect the mean weight of all fish caught to be higher or lower than the
median? Explain.
b. You caught 3 bass weighing 2.3 pounds, 3.9 pounds, and 4.2 pounds. Were any of your
fish outliers (according to the original statistics above)? Explain.

Answer: a. Probably higher. The data appear to be skewed to the right.
b. IQR = 3.3 - 2.8 = 0.5. Since 1.5(IQR) = 0.75, the fences are 2.8 - 0.75 = 2.05 and 3.3
+ 0.75 = 4.05. The fish weighing 4.2 pounds is more than 1.5 IQRs outside the
quartiles, so it could be considered an outlier.

47) The boxplots show prices of used cars (in thousands of dollars) advertised for sale at three different car dealers.
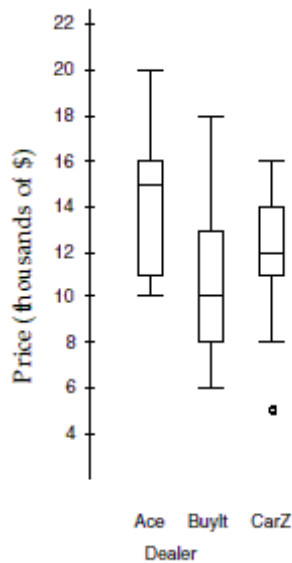
a. Which dealer offers the cheapest car offered, and at what price?
b. Which dealer has the lowest median price, and how much is it?
c. Which dealer has the smallest price range, and what is it?
d. Which dealer's prices have the smallest IQR, and what is it?
e. Which dealer generally sells cars cheapest? Explain.

Answer: a. Car Z:$5000
      b. BuyIt:$10,000
      c. Ace: $10,000
      d. CarZ:$3000
      e. BuyIt; half of their cars are cheaper than any of the cars at Ace, and 25% of their cars are cheaper than all but one car at CarZ. The third quartile of their prices is well below the third quartile at CarZ, and below even the median price at Ace.

47) _____

48) At www.census.gov you can create a "population pyramid" for any country. These   48) _____
pyramids are back-to-back histograms. This pyramid shows Mexico's 2000 female
population and the census bureau's projection for 2050. Write a few sentences
summarizing the changes that are forecast.

Source: U.S. Census Bureau, International Data Base.
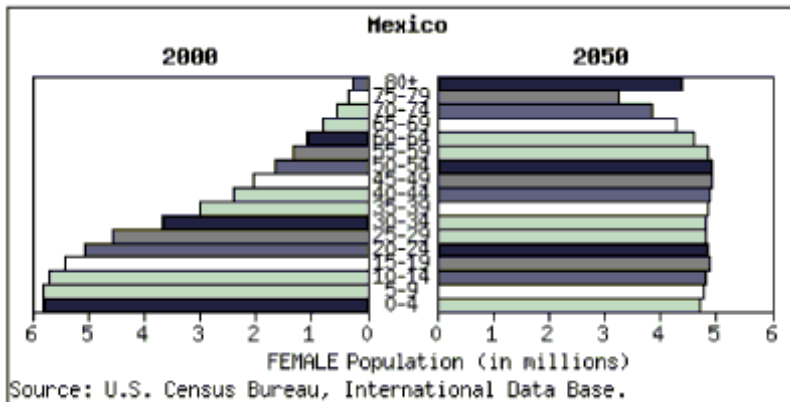
Answer: The Census Bureau projects dramatic changes in the female population of Mexico
over the next 50 years. The current distribution of ages is strongly skewed to the
right with most of the women under 30 and far fewer 50 and above. By 2050 the
population will become more uniform across age groups from 0 to 60, and we
anticipate an unusually large number of women over 80.

49) The five-number summary for the fuel economy (in miles per gallon) of year 2011 midsize   49) _____
cars is:

Min	Q1	Median	Q3	Max
13	21	23	26	50

a. Would you expect the mean gas mileage of all midsize cars to be higher or lower than
the median? Explain.
b. One model of Volkswagon gets 34 mpg, one model of Toyota gets 28 mpg, and one
model of Bentley gets 13 mpg. Are any of these cars outliers (according to the original
statistics above)? Explain.

Answer: a. Probably higher. The data appear to be skewed to the right.
b. IQR = 26 – 21 = 5. Since 1.5(IQR) = 7.5, the fences are 21 – 7.5 = 13.5 and 26 + 7.5 =
33.5. The Volkswagon with 34 mpg is more than 1.5 IQRs above Q3 and the Bentley
with 13 mpg is more than 1.5 IQR below Q1, so they could both be considered
outliers.

50) The boxplots show fuel economy of 2011 model cars for the classes shown.   50) _____
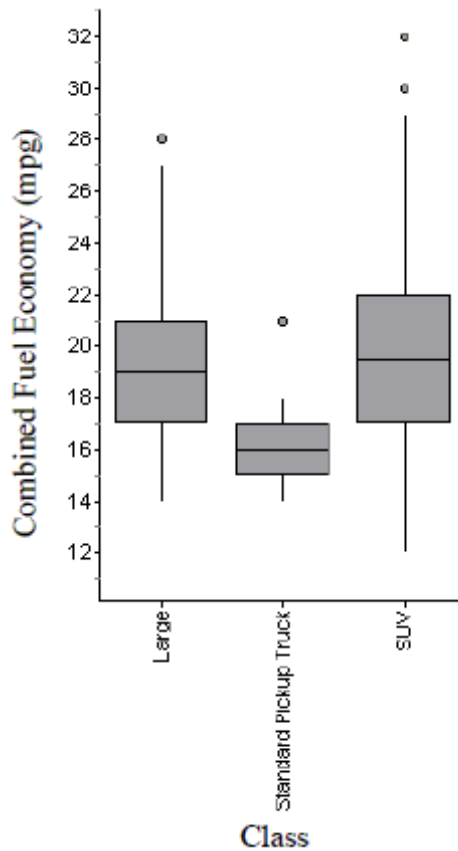
a. Which class offers the car with the best gas mileage, and what is the mileage of that car?
b. Which class has the highest median gas mileage, and how much is it?
c. Which class of car has the smallest range of fuel economy, and what is it?
d. Which class of car has the smallest IQR, and what is it?
e. Which class of car generally gets the best mileage? Explain.

Answer: a. SUV: 32 mpg
          b. SUV: 19.5 mpg
          c. Pickup: 7 mpg
          d. Pickup: 2 mpg

          e. SUV; $\frac{3}{4}$ of SUVs get better mileage than $\frac{3}{4}$ of the Pickup Trucks. SUVs have the

          highest median, third quartile, and maximum. The first quartile is the same as that
          of Large cars, and the minimum is the lowest of all three, but overall SUVs appear to
          have the best mileage.

51) How do sports salaries compare? Two sets of histograms below show the distributions of   51) _____
salaries for Major League Baseball and the National Football League. Give two reasons
why the second set of histograms makes it easier to compare the distributions. Then write a
few sentences comparing the salary distributions for the two sports.

Histogram set 1

2013 Salary (millions of $)

Histogram set 2

2013 Salary (millions of $)
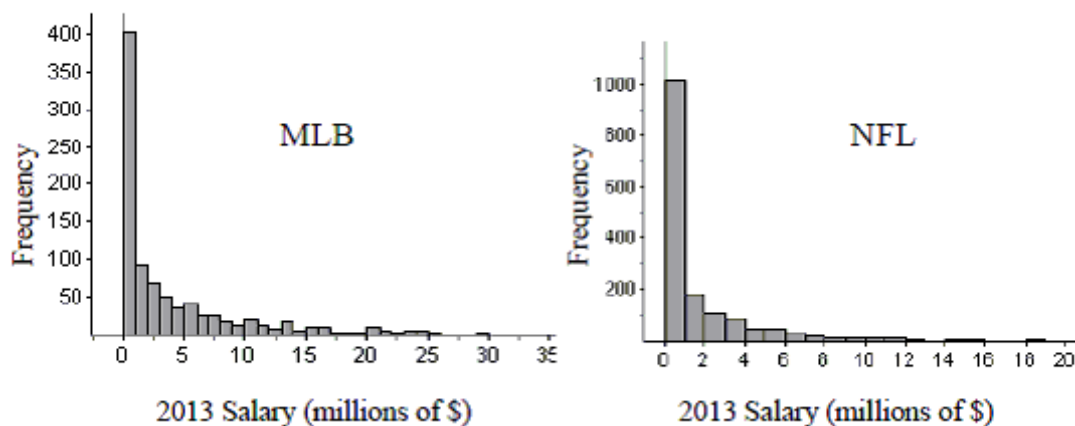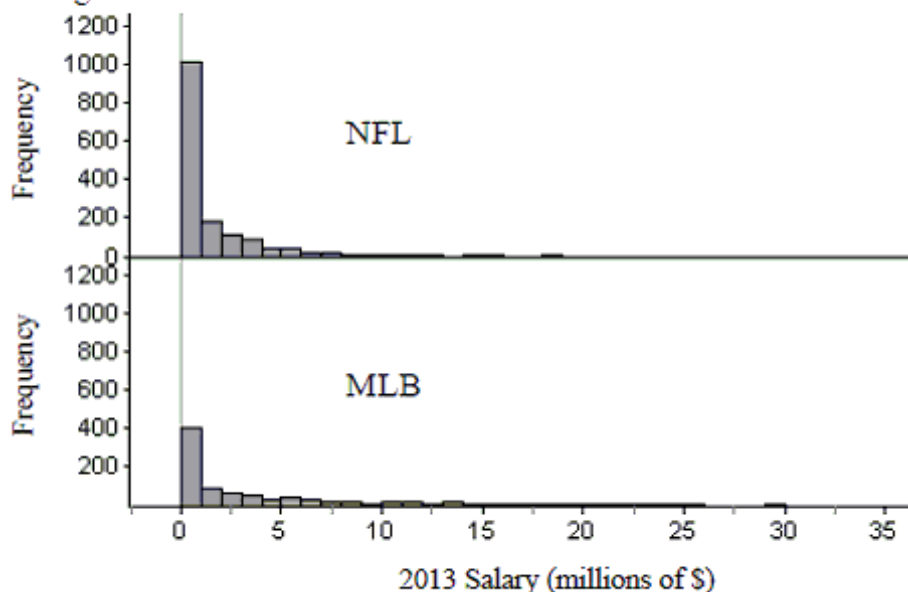
Answer: The second set of histograms is stacked vertically, and they use the same scales on
the axes. Both distributions are strongly skewed to the right. MLB salaries are higher
(the median is higher, and there are fewer low salaries and the highest salaries are
much higher.) The MLB salaries also have much more variability (greater spread).

52) Students taking an intro stats class reported the number of credit hours that they were taking that quarter. Summary statistics are shown in the table.

$\bar{x}$	16.65
s	2.96
min	5
Q1	15
median	16
Q3	19
max	28

a. Suppose that the college charges $73 per credit hour plus a flat student fee of $35 per quarter. For example, a student taking 12 credit hours would pay $35 + $73(12) = $911 for that quarter.
  i. What is the mean fee paid?
  ii. What is the standard deviation for the fees paid?
  iii. What is the median fee paid?
  iv. What is the IQR for the fees paid?
b. Twenty-eight credit hours seems like a lot. Would you consider 28 credit hours to be unusually high? Explain.

Answer: a. i. $35 + $73(16.65) = $1250.45
        ii. $73(2.96) = $216.08
        iii. $35 + $73(16) = $1203
        iv. IQR = 73(19 - 15) = $292
        b. IQR = 19 - 15 = 4 credit hours
        High outliers will lie above Q3 + 1.5IQR = 19 + 1.5(4) = 25 credit hours. Since 28 credit hours exceeds 25 credit hours, I would consider 28 credit hours to be unusually high.

53) The Wechsler Adult Intelligence Scale - Revised (WAIS-R) follows a Normal model with mean 100 and standard deviation 15. Draw and clearly label this model.
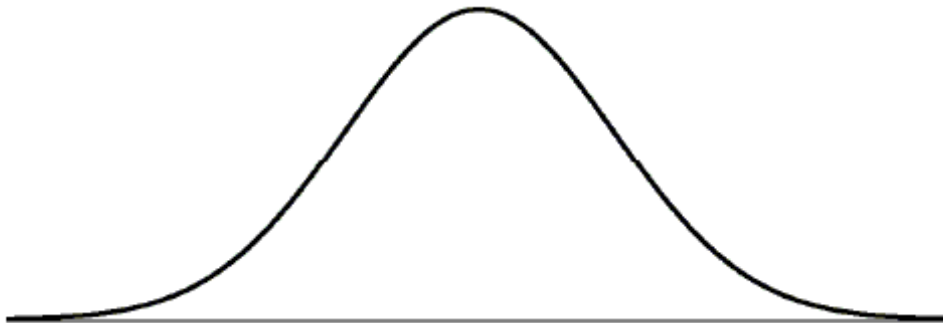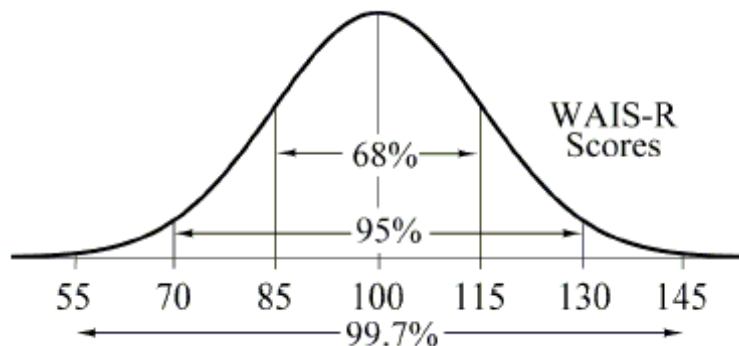
Answer:

54) Adult female Dalmatians weigh an average of 50 pounds with a standard deviation of 3.3 pounds. Adult female Boxers weigh an average of 57.5 pounds with a standard deviation of 1.7 pounds. One statistics teacher owns an underweight Dalmatian and an underweight Boxer. The Dalmatian weighs 45 pounds, and the Boxer weighs 52 pounds. Which dog is more underweight? Explain.

54) _____

Answer: Dalmatian: $z_D = \dfrac{45 - 50}{3.3} = -1.52$

Boxer: $z_B = \dfrac{52 - 57.5}{1.7} = -3.24$

The Dalmatian is 1.52 standard deviations underweight, while the Boxer is 3.24 standard deviations underweight. So, the Boxer is more underweight.

55) Human body temperatures taken through the ear are typically 0.5° F higher than body temperatures taken orally. Making this adjustment and using the 1992 *Journal of the American Medical Association* article that reports average oral body temperature as 98.2° F, we will assume that a Normal model with an average of 98.7° F and a standard deviation of 0.7° F is appropriate for body temperatures taken through the ear.
a. An ear temperature of 97° F may indicate hypothermia (low body temperature). What percent of people have ear temperatures that may indicate hypothermia?
b. Find the interquartile range for ear temperatures.
c. A new thermometer for the ear reports that it is more accurate than the ear thermometers currently on the market. If the average ear temperature reading remains the same and the company reports an IQR of 0.5° F, find the standard deviation for this new ear thermometer.

55) _____

Answer: a. $z = \dfrac{97 - 98.7}{0.7} = -2.43$, so $P(z < -2.43) = 0.0075$

About 0.75% of people have ear temperatures that may indicate hypothermia.
b. The z-scores associated with the IQR are $z = -0.67$ and $z = 0.67$. So, we need to solve for y in each of the following equations: $-0.67 = \dfrac{y - 98.7}{1.7}$ and $0.67 = \dfrac{y - 98.7}{1.7}$.

We get $y = 98.7 - 0.67(0.7) = 98.2$ and $y = 98.7 + 0.67(0.7) = 99.2$. The interquartile range is IQR = 99.2° F - 98.2° F = 1.0° F.
c. The new IQR is 0.5° F, while the old IQR was 1.0° F. So, we want

IQR = $[98.7 + 0.67\sigma] - [98.7 - 0.67\sigma] = 0.5$, or $1.34\sigma = 0.5$. Thus, $\sigma = \dfrac{0.5}{1.34} = 0.37$. Our

new standard deviation is 0.37° F.

56) During a budget meeting, local school board members decided to review class size   56) _____
information to determine if budgets were correct. Summary statistics are shown in the
table.

$\bar{x}$	33.39 students
s	5.66 students
min	17
Q1	29
median	33
Q3	40
max	40

a. Notice that the third quartile and maximum class sizes are the same. Explain how this
can be.
b. The school district declares that classes with enrollments fewer than 20 students are "too
small". Would you consider a class of 20 students to be unusually small? Explain.
c. The school district sets the office supply budgets of their high schools on the enrollment
of students. The district budgets each class $12 plus $0.75 per student, so a class with one
student receives $12.75 and the classes with 40 students receive 12 + 0.75(40) = $42. What is
the median class budget for office supplies? And the IQR?
d. What are the mean and standard deviation of the class office supply budgets?

Answer: a. The top 25 percent of all classes have 40 students enrolled.
b. Yes, classes with 20 students enrolled seem unusually small. Twenty is well below
the first
quartile of 29 students, and only slightly above the minimum size (17). z =
$\dfrac{20 - 33.39}{5.66}$ = -2.366 With z = -2.366, this size class is over 2 standard deviation units

below the mean.
c. Median budget = $12 + $0.75(33) = $36.75
Q1 budget = $12 + $0.75(29) = $33.75
Q3 budget = $12 + $0.75(40) = $42.00
IQR = $42.00 - $33.75 = $8.25
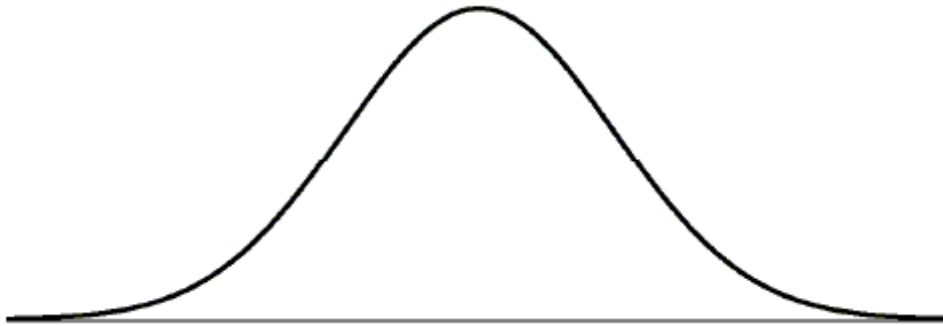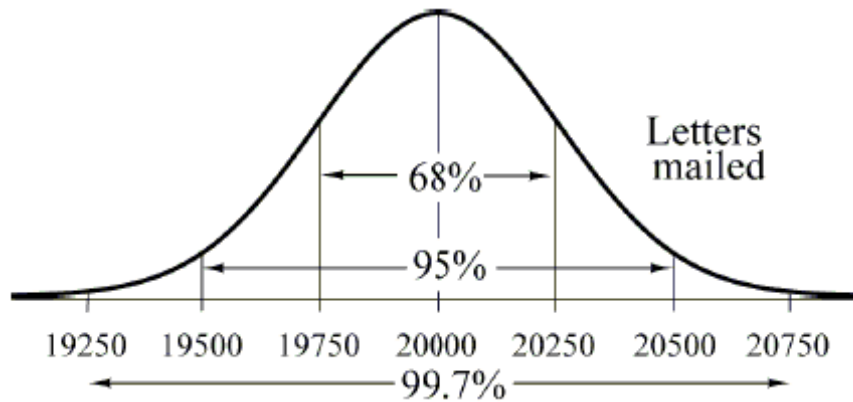d. Mean budget = $12 + $0.75(33.39) = $37.04
Standard deviation = $0.75(5.66) = $4.25

57) The Postmaster of a city's Post Office believes that a Normal model is useful in projecting the number of letters which will be mailed during the day. They use a mean of 20,000 letters and a standard deviation of 250 letters. Draw and clearly label this model.

57) _____

Answer:

58) Light bulbs are measured in lumens (light output), watts (energy used), and hours (life). A standard white light bulb has a mean life of 675 hours and a standard deviation of 50 hours. A soft white light bulb has a mean life of 700 hours and a standard deviation of 35 hours. At a local science competition, both light bulbs lasted 750 hours. Which light bulb's life span was better? Explain.

58) _____

Answer: Standard light bulb: $z = \dfrac{750 - 675}{50} = 1.5$

Soft light bulb: $z = \dfrac{750 - 700}{35} = 1.4286$

The standard light bulb lasted more than 1.5 standard deviations above the mean life, compared to the soft light bulb at 1.4286 standard deviations above its mean. The standard light bulb's performance was slightly better.

59) At a large business, employees must report to work at 7:30 A.M. The arrival times of employees can be described by a Normal model with mean of 7:22 A.M. and a standard deviation of 4 minutes.
(SHOW WORK)
a. What percent of employees are late on a typical work day?
b. A psychological study determined that the typical worker needs five minutes to adjust to their surroundings before beginning their duties. What percent of this business' employees arrive early enough to make this adjustment?
c. Because late employees are a distraction and cost companies money, all employees need to be on time to work. If the mean arrival time of employees does not change, what standard deviation would the arrival times need to ensure virtually all employees are on time to work?

59) _____

25

d. Explain what achieving a smaller standard deviation means in the context of this problem.
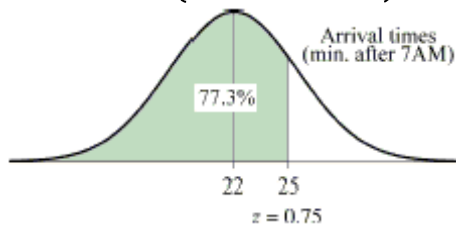
Answer: a. Employees are late if they arrive after 7:30 AM..

$$P(\text{time} > 7:30) = P\left(z > \frac{30 - 22}{4} = 2\right) = 0.0228$$
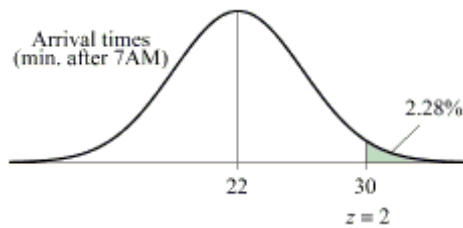
According to the Normal model, about 2.28% of employees are expected to arrive after 7:30 AM.

b. $P(x \le 25) = P\left(z < \frac{25 - 22}{4} = 0.75\right) = 0.773$

According to the Normal model, about 77.3% of employees arrive at work before 7:25 AM.

c. Virtually all times lie within 3 standard deviations of the mean. (Accept other reasonable z-scores greater than 3). If $z = 3 \le \frac{30 - 22}{\sigma}$, then $3\sigma \le 8$, so $\sigma \le 2.67$ minutes.

d. A smaller standard deviation would mean greater consistency in arrival times.

60) City planners wanted to know how many people lived in a typical housing unit so they compiled data from hundreds of forms that had been submitted in various city offices. Summary statistics are shown in the table.

| $\bar{x}$	2.53 people
s	1.40 people
min	1
Q1	1
median	2
Q3	3
max	10

a. Notice that the minimum occupancy and the first quartile are the same. Explain how this can be.

b. The city classifies residences housing 4 or more people as "high occupancy". Would you consider 4 occupants to be unusually high? Explain.

c. The city bases their garbage disposal fee on the occupancy level of the home or apartment. The annual fee is $40 plus $5 per person, so a single occupant pays $45 and the homes with 10 people pay $40 +$5(10) = $90 a year. What is the median fee paid? And the IQR?

d. What are the mean and standard deviation of the garbage disposal fees?

Answer: a. At least 25% of the housing units have only one resident.

b. 4 residents is just above the 3rd quartile, and only about one standard deviation above the mean. It is not an unusually high number of residents.

c. Median = 40 + 5(2) = $50; IQR = 5(3 - 1) = $10

d. Mean = 40 + 5(2.53) = $52.65; SD = 5(1.40) = $7.00
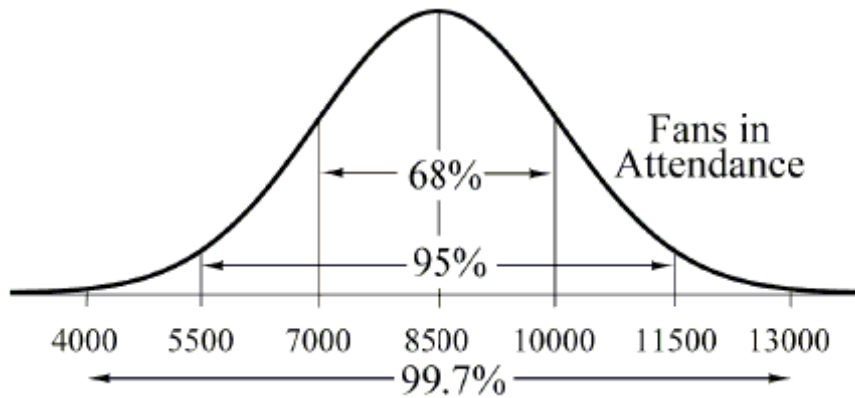
61) Owners of a minor league baseball team believe that a Normal model is useful in projecting the number of fans who will attend home games. They use a mean of 8500 fans and a standard deviation of 1500 fans. Draw and clearly label this model.

61) _____
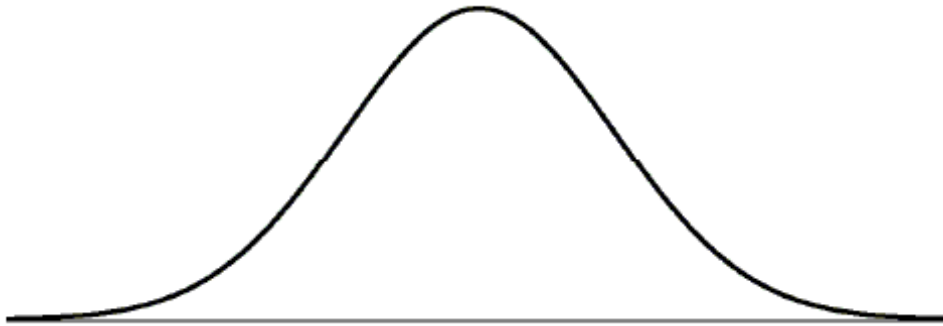
Answer:

62) Although most of us buy milk by the quart or gallon, farmers measure daily production in pounds. Guernsey cows average 39 pounds of milk a day with a standard deviation of 8 pounds. For Jerseys the mean daily production is 43 pounds with a standard deviation of 5 pounds. When being shown at a state fair a champion Guernsey and a champion Jersey each gave 54 pounds of milk. Which cow's milk production was more remarkable? Explain.

62) _____

Answer: The Jersey's milk production was comparatively higher. That cow gave slightly more than 2 standard deviations above the average amount of milk ($z = 2.2$), while the Guernsey gave less than 2 standard deviations more than the average for Guernseys ($z = 1.875$).

63) A company's manufacturing process uses 500 gallons of water at a time. A "scrubbing" machine then removes most of a chemical pollutant before pumping the water into a nearby lake. Legally the treated water should contain no more than 80 parts per million of the chemical, but the machine isn't perfect and it is costly to operate. Since there's a fine if the discharged water exceeds the legal maximum, the company sets the machine to attain an average of 75 ppm for the batches of water treated. They believe the machine's output can be described by a Normal model with standard deviation 4.2 ppm. (SHOW WORK)

63) _____

a. What percent of the batches of water discharged exceed the 80ppm standard?
b. The company's lawyers insist that they not have more than 2% of the water over the limit. To what mean value should the company set the scrubbing machine? Assume the standard deviation does not change.
c. Because achieving a mean that low would raise the costs too much, they decide to leave the mean set at 75 ppm and try to reduce the standard deviation to achieve the "only 2%

28

over" goal. Find the new standard deviation needed.

d. Explain what achieving a smaller standard deviation means in this context.

Answer: a. According to the normal model, we expect about 11.7% of the batches to exceed the 80 ppm standard.

Chemical level (ppm)

11.7%

75   80

$z = 1.19$
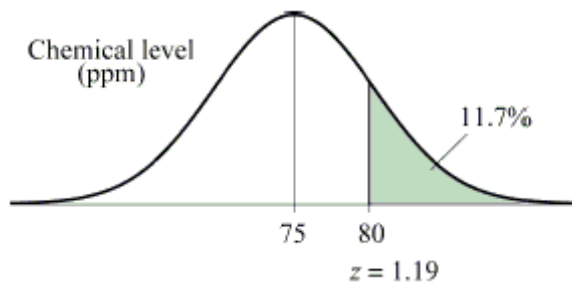
b. According to the Normal model, a mean of about 71.37ppm would need to be achieved.

Chemical level (ppm)

2%

μ 80

$z = 2.054$
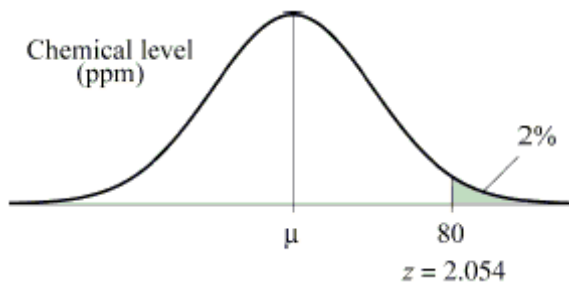
c. According to the Normal model, the new standard deviation would need to be at most 2.43ppm.

d. The scrubber must be more consistent in its performance from batch to batch.

64) A statistics teacher gave her class a 15 point quiz. The summary statistics for the students'   64) _____
scores are shown in the table.

$\bar{x}$	10.95 points
s	2.481 points
min	4
Q1	9.5
median	12
Q3	12
max	15

a. Notice that the median score and the third quartile are the same. Explain how this can be.

b. One student's parent heaped praise on him for scoring 13, saying it was an amazing score. Comment on whether that praise is deserved using the summary statistics as support.

c. To convert these raw scores to a score out of 100, the teacher multiplies each score by six, then adds 10. (We can debate the wisdom of such a strategy later!). What is the median converted score? And the IQR?

d. What are the mean and standard deviation of the converted test scores?

Answer: a. At least 25% of the students units had a score of 12.

b. The parent might be saying that the student did well compared to his usual performance, but as the class scores go it's not exceptionally strong. While we can say it was better than 75% of the students, we can't say anything stronger than that.

c. Median = 6(12) + 10 = 82 points; IQR = 6(12 – 9.5) = 15 points

d. Mean = 6(10.95) + 10 = 75.7 points; SD = 6(2.481) = 14.866 points

65) Nickels minted in the United States are supposed to weigh 5.000 grams. Of course there is some variation in that. The actual weights are pretty well represented by a normal model with a mean of 5.000 g and a standard deviation of about 0.08 g. Draw and clearly label this model.

65) _____

Answer:
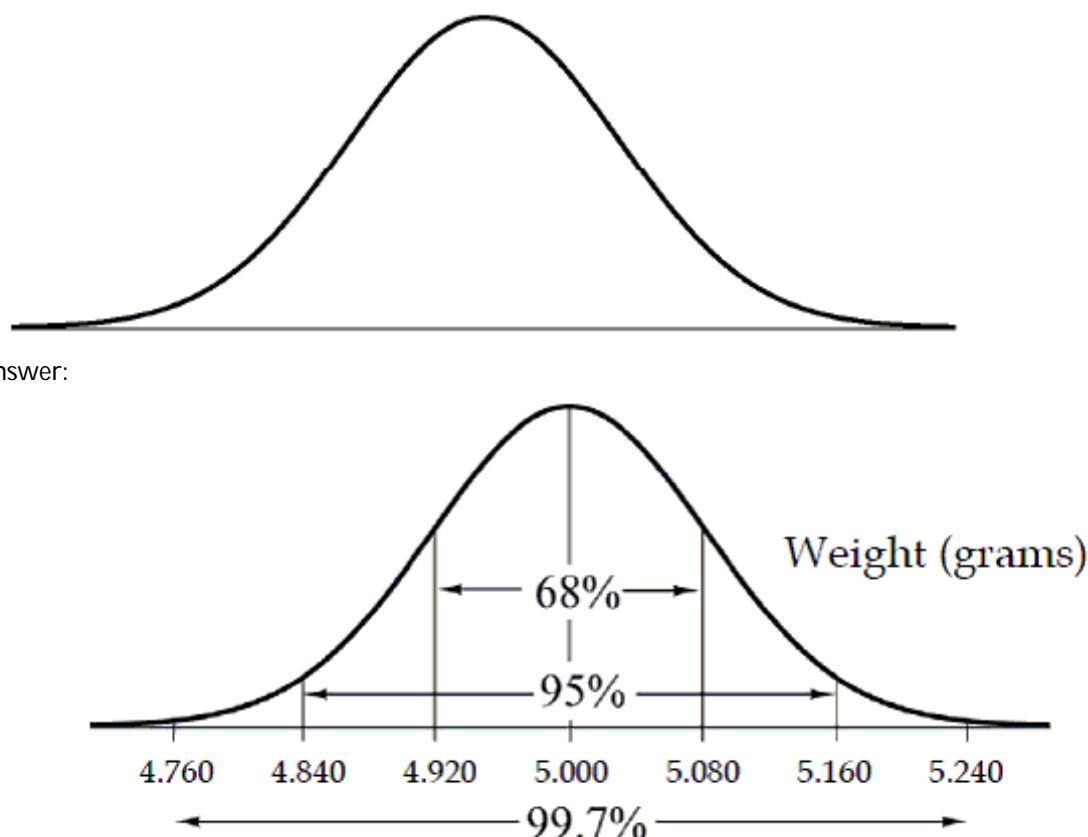
66) Dimes minted in the United States average 2.286 g with a standard deviation of 0.06 g. Nickels have a mean of 5.000 g and a standard deviation of about 0.08 g. A couple chemistry students were trying out their teacher's new scale by weighing a bunch of coins. They found a nickel that weighed 5.19 g and a dime that weighed 2.45 g. Which coin was more exceptionally heavy? Explain.

66) _____

Answer: The dime is more unusual. The nickel weighed a bit more than 2 standard deviations above the average weight ($z = 2.375$), while the dime was closer to 3 standard deviations more than the average for dimes ($z = 2.733$).

67) The lifespans of a particular brand of graphing calculator are approximately normally distributed with a mean of 620 days from the purchase date and a standard deviation of 82 days. They will provide a warranty that guarantees a replacement if the calculator stops working within the specified time frame, and are trying to decide what time frame to use.
a. If the company sets the warranty at a year and a half (say 540 days), what proportion of calculators will they have to replace?
b. The company does not want to have to replace more than 1% of the calculators they sell. What length of time should they set for the warranty?
c. The company would like to set the warranty for 540 days, and still replace no more than 1% of the calculators sold. Increasing the average life of the calculators is too expensive, but they think they reduce the standard deviation of the lifespans. What standard deviation of lifespans would be needed to make this happen?
d. Explain what achieving a smaller standard deviation means in this context.

67) _____

Answer: a. According to the normal model, we expect about 16.45% of the calculators to break down before 540 days.

z = -0.976

Days since purchase

540   620
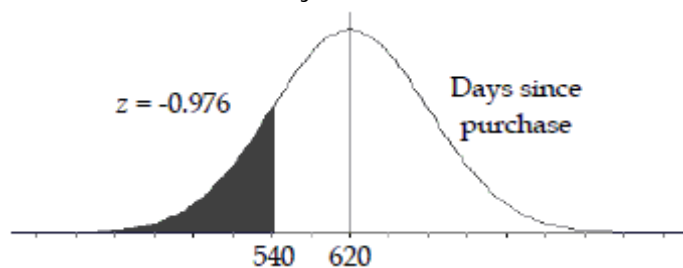
b. According to the normal model, they would need to set the warranty at 429 days.

z = -2.326

Days since purchase

429   620
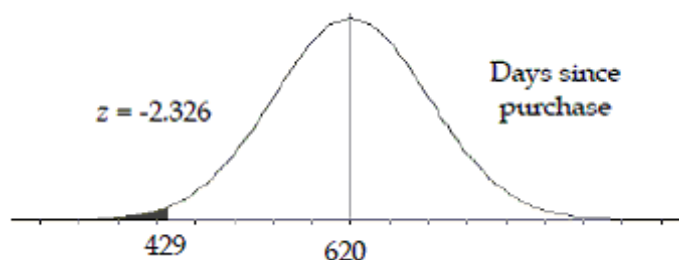
c. According to the normal model, 540 must be 2.326 standard deviations below the mean, so the standard deviation is about 34.4 days.

z = -2.326

Days since purchase

540   620
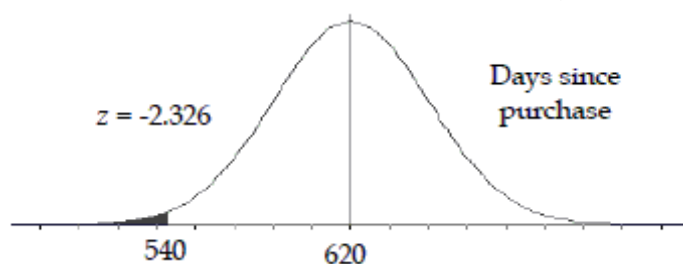
d. It means they would have to make the calculators more consistent in their lifespan.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

68) School administrators collect data on students attending the school. Which of the following variables is quantitative?
   A) grade point average
   B) class (freshman, soph., junior, senior)
   C) whether the student is in AP* classes
   D) none of these
   E) whether the student has taken the SAT

68) _____

Answer: A

69) Which of the following variables would most likely follow a Normal model?
   A) all of these
   B) family income
   C) scores on an easy test
   D) weights of adult male elephants
   E) heights of singers in a co-ed choir

69) _____

Answer: D

70) A professor has kept records on grades that students have earned in his class. If he wants to examine the percentage of students earning the grades A, B, C, D, and F during the most recent term, which kind of plot could he make?

    A) boxplot   B) histogram   C) timeplot   D) pie chart   E) dotplot

70) _____

Answer: D

71) Which is true of the data shown in the histogram?

71) _____
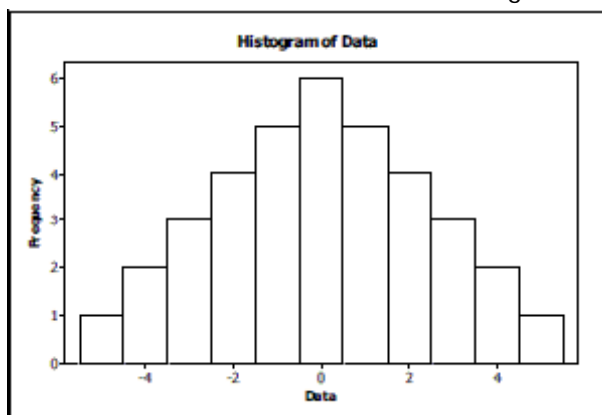
I. The distribution is approximately symmetric.
II. The mean and median are approximately equal.
III. The median and IQR summarize the data better than the mean and standard deviation.

    A) III only
    B) I only
    C) I and III
    D) I and II
    E) I, II, and III

Answer: D

72) Two sections of a class took the same quiz. Section A had 15 students who had a mean score of 80, and Section B had 20 students who had a mean score of 90. Overall, what was the approximate mean score for all of the students on the quiz?

72) _____

    A) 85.0
    B) none of these
    C) 84.3
    D) 85.7
    E) It cannot be determined.

Answer: D

73) Your Stats teacher tells you your test score was the 3rd quartile for the class. Which is true?
I. You got 75% on the test.
II. You can't really tell what this means without knowing the standard deviation.
III. You can't really tell what this means unless the class distribution is nearly Normal.

73) _____

    A) III only
    B) II and III
    C) II only
    D) I only
    E) none of these

Answer: E