

**Chapter 1:**

1. Graphs: Be able to read and create dotplot, stemplot, boxplot & histogram

- Be sure that all of your graphs include Scales & labels !!!

2. What are the measures of center? Of spread?

Center mean median      Spread Standard deviation IQR

3. How do you describe a distribution:

C center      Boxplots use median and IQR  
 U unusual features  
 S Shape      All other graphs - mean and std. dev  
 S Spread

4. How do you calculate outliers?

Any data point outside  $Q_1 - 1.5 IQR$  or  $Q_3 + 1.5 IQR$

5. Which statistics are resistant?

Median, IQR

6. What are frequency, relative frequency & cumulative frequency?

counts      %      sum of counts  
 $\frac{\text{count}}{\text{total}}$

7. What are percentiles?

The  $n$ th percentile signifies  $n\%$  of the data falls at or below the given value.

8. The following statistics were calculated from the distribution of a set of scores:

Mean	Median	Range	IQR	Std Dev	Variance
3	3	4	3	1.414	2

WITHOUT using a calculator, complete the following:

a) If 7 is added to each of the original scores, what will the statistics be for the distribution of the new scores?

Mean	Median	Range	IQR	Std Dev	Variance
10	10	4	3	1.414	2

b) If each of the original scores is multiplied by 4, what will the statistics be for the distribution of the new scores?

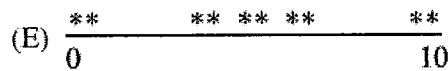
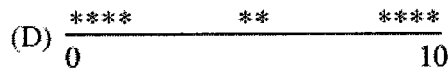
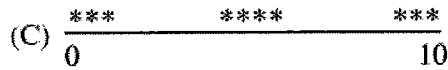
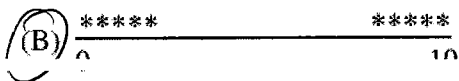
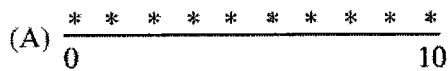
Mean	Median	Range	IQR	Std Dev	Variance
12	12	16	12	5.656	32

c) If the original scores are represented by  $X$ , what will the statistics be for the distribution of  $4X + 7$ ?

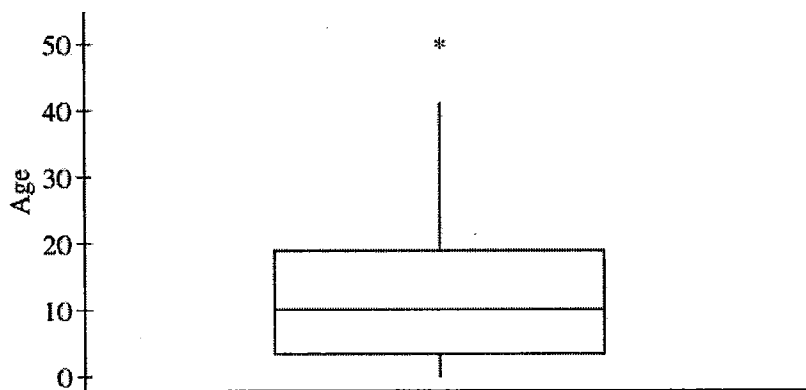
Mean	Median	Range	IQR	Std Dev	Variance
19	19	16	12	5.656	32

Center + 7  
 All x 4  
 variance x 16  
 All x 4  
 var x 16  
 center + 7

9. Of the following dotplots, which represents the set of data that has the greatest standard deviation?

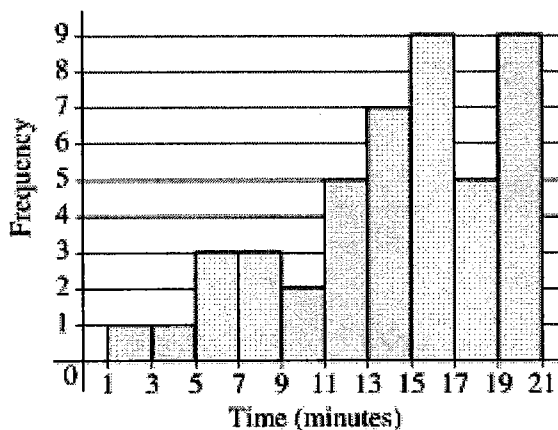


10. A random sample of 374 United States pennies was collected, and the age of each penny was determined. According to the boxplot below, what is the approximate interquartile range (IQR) of the ages?



- (A) 8      (B) 10      (C) 16      (D) 40      (E) 50

11. The histogram shows the number of minutes needed by 45 students to finish playing a computer game. Which of the following statements is correct?



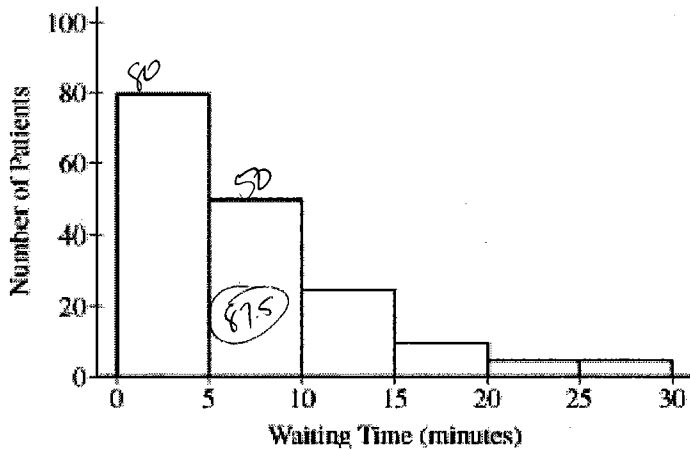
- (A) The distribution is skewed to the right.  
 (B) The distribution is skewed to the left.  
 (C) The distribution appears to be normal.  
 (D) The distribution appears to be chi-square.  
 (E) The distribution appears to be uniform.

12. A random sample of size 10 was taken from a population. The sample has a variance of zero. Which of the following statements must be true?

- I. The population also has a variance of zero.
- II. The sample mean is equal to the sample median.
- III. The ten data points in the sample are equal in numerical value.

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

13. The histogram below displays the frequencies of waiting times, in minutes, for 175 patients in a dentist's office.



Median = patient 87.5  
between 5: 10 min.

Which of the following could be the median of the waiting times, in minutes?

(A) 2.50                      (B) 7.25                      (C) 12.25                      (D) 15.00                      (E) 17.50

14. Data were collected on the amount, in dollars, that individual customers spent on dinner in an Italian restaurant. The quartiles for these data are given below.

Q1	Q2	Q3
\$36.27	\$44.27	\$58.97

Which of the following statements must be true for these customers?

- (A) At least half of the customers spent less than or equal to \$44.27 and at least half spent greater than or equal to \$44.27.
- (B) Seventy-five percent of the customers spent between \$36.27 and \$58.97.
- (C) Twenty-five percent of the customers spent less than or equal to \$58.97 and the remaining 75 percent spent greater than or equal to \$58.97.
- (D) The mean amount spent by customers is \$44.27. *not known*
- (E) A majority of customers spent \$44.27

15. A company sells concrete in batches of 5 cubic yards. The probability distribution of  $X$ , the number of cubic yards sold in a single order for concrete from this company, is shown in the table below.

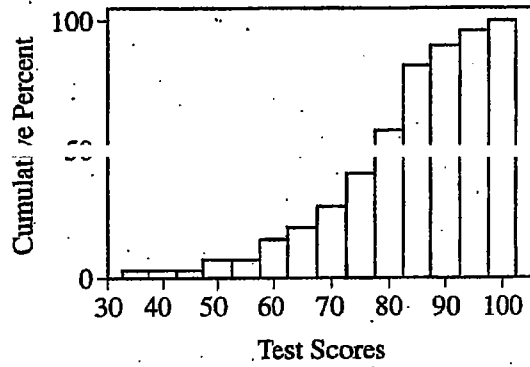
$X =$ the number of cubic yards	10	15	20	25	30
Probability	0.15	0.25	0.25	0.30	0.05

75(5.76)

The expected value of the probability distribution of  $X$  is 19.25 and the standard deviation is 5.76. There is a fixed cost to deliver the concrete. The profit  $Y$ , in dollars, for a particular order can be described by  $Y = 75X - 100$ . What is the standard deviation of  $Y$ ?

(A) \$332.00                      (B) \$432.00                      (C) \$532.00                      (D) \$1,343.75                      (E) \$1,400.00

AP STATISTICS  
TEST SCORES

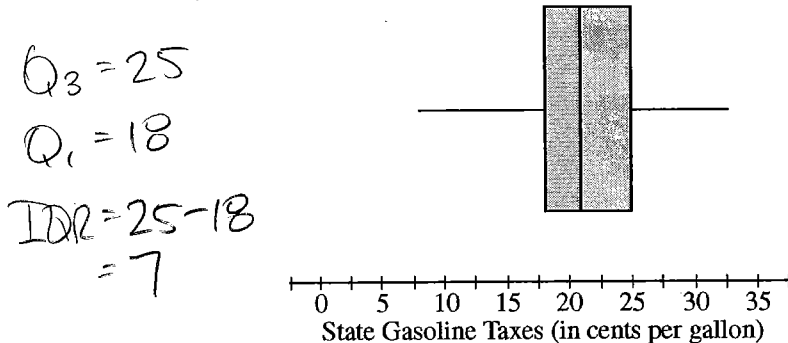


16. The figure above shows a cumulative relative frequency histogram of 40 scores on a test given in an AP Statistics class. Which of the following conclusions can be made from the graph?

- (A) There is greater variability in the lower 20 test scores than in the higher 20 test scores.
- (B) The median test score is less than 50.
- (C) Sixty percent of the students had test scores above 80.
- (D) If the passing score is 70, most students did not pass the test.
- (E) The horizontal nature of the graph for test scores of 60 and below indicates that those scores occurred most frequently.

17. (2009B #1)

As gasoline prices have increased in recent years, many drivers have expressed concern about the taxes they pay on gasoline for their cars. In the United States, gasoline taxes are imposed by both the federal government and by individual states. The boxplot below shows the distribution of the state gasoline taxes, in cents per gallon, for all 50 states on January 1, 2006.



(a) Based on the boxplot, what are the approximate values of the median and the interquartile range of the distribution of state gasoline taxes, in cents per gallon? Mark and label the boxplot to indicate how you found the approximated values.

Median  $\approx$  21¢/gal    IQR = 7¢/gal

(b) The federal tax imposed on gasoline was 18.4 cents per gallon at the time the state taxes were in effect. The federal gasoline tax was added to the state gasoline tax for each state to create a new distribution of combined gasoline taxes.

What are approximate values, in cents per gallon, of the median and interquartile range of the new distribution of combined gasoline taxes? Justify your answer.

Median =  $21 + 18.4 = 39.4$  ¢/gal  
IQR remains 7¢/gal

Recentering only affects center, not spread

18. (2008B #1)

A certain state's education commissioner released a new report card for all the public schools in that state. This report card provides a new tool for comparing schools across the state. One of the key measures that can be computed from the report card is the student-to-teacher ratio, which is the number of students enrolled in a given school divided by the number of teachers at that school. The data below give the student-to-teacher ratio at the 10 schools with the highest proportion of students meeting the state reading standards in the third grade and at the 10 schools with the lowest proportion of students meeting the state reading standards in the third grade.

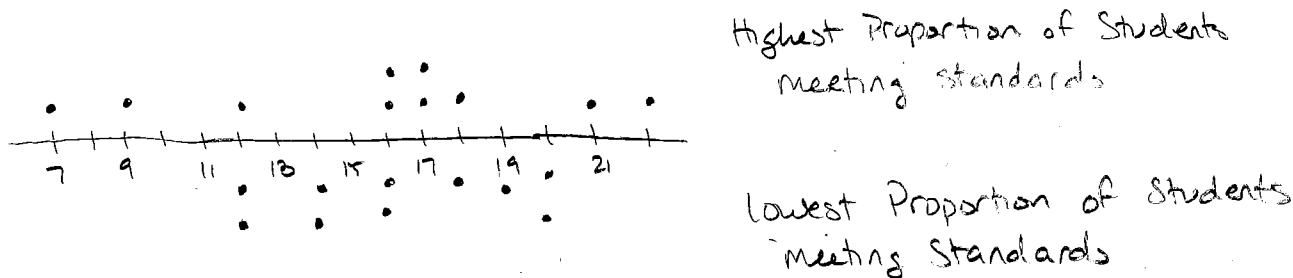
Ratios in the 10 Schools with Highest Proportion of Students Meeting Standards

7 21 18 22 9 16 12 17 17 16

Ratios in the 10 Schools with Lowest Proportion of Students Meeting Standards

14 16 18 20 12 14 16 12 20 19

(a) Display a dotplot for each group to compare the distribution of student-to-teacher ratios in the top 10 schools with the distribution in the bottom 10 schools. Comment on the similarities and differences between the two distributions.



miss  
Both have approximately the same center with high proportion schools at 16.5 and low proportion schools at 16. Neither have unusual features. High proportion schools are approximately normal while low proportion schools are uniform. Spreads are also different with high proportion schools having a larger standard deviation than low proportion schools.

(b) Any statistical test that is used to determine whether the mean student-to-teacher ratio is the same for the top 10 schools as it is for the bottom 10 schools would be inappropriate. Explain why in a few sentences.

This is not a simple random sample of schools. Data is taken from an ordered set of schools, not a random sample of the population.

## Chapter 2:

19. What are z-scores and how do you calculate them?

$$z = \frac{x - \mu}{\sigma}$$

Measures the number of standard deviations from the mean

20. What is the mean of a density curve? The median?

Mean - balances the curve

Median - divides the curve into two equal areas

21. What is the empirical rule?

Generalization for the normal curve where about 68% of the data falls within one standard deviation from the mean, 95% lies within two standard deviations and 99.7% lie within 3 std. dev.

22. When would you use a normal probability plot? (See p. 150)

The plot should be linear to check a distribution for normality.

23. The weight of adult male grizzly bears living in the wild in the continental United States is approximately normally distributed with a mean of 500 pounds and a standard deviation of 50 pounds. The weight of adult female grizzly bears is approximately normally distributed with a mean of 300 pounds and a standard deviation of 40 pounds. Approximately, what would be the weight of a female grizzly bear with the same standardized score (z-score) as a male grizzly bear with a weight of 530 pounds?  $z = \frac{530 - 500}{50} = .6$   $.6 = \frac{x - 300}{40}$   $x = 324$

- (A) 276 pounds    (B) 324 pounds    (C) 330 pounds    (D) 340 pounds    (E) 530 pounds

24. The commuting time for a student to travel from home to a college campus is normally distributed with a mean of 30 minutes and a standard deviation of 5 minutes. If the student leaves home at 8:25 A.M., what is the probability that the student will arrive at the college campus later than 9 A.M.?

- (A) 0.16    (B) 0.32    (C) 0.50    (D) 0.84    (E) 1.00

$$z = \frac{35 - 30}{5} = 1 \quad P(z > 1) = .159$$

25. There is no #25! ☹

26. The distribution of the weights of loaves of bread from a certain bakery follows an approximately normal distribution. Based on a very large sample, it was found that 10 percent of the loaves weighed less than 15.34 ounces, and 20 percent of the loaves weighed more than 16.31 ounces. What are the mean and standard deviation of the distribution of the weights of the loaves of bread?

- (A)  $\mu = 15.82, \sigma = 0.48$   
 (B)  $\mu = 15.82, \sigma = 0.69$   
 (C)  $\mu = 15.87, \sigma = 0.50$   
 (D)  $\mu = 15.93, \sigma = 0.46$   
 (E)  $\mu = 16.00, \sigma = 0.50$

$$z^*(.10) = -1.282 \quad -1.282 = \frac{15.34 - \mu}{\sigma} \quad .842 = \frac{16.31 - \mu}{\sigma}$$

$$z^*(.80) = .842$$

$$-1.282\sigma = 15.34 - \mu \quad .842\sigma = 16.31 - \mu$$

$$-1.282\sigma = 15.34 - \mu$$

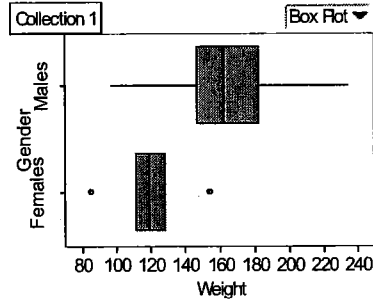
$$-1.842\sigma = 16.31 - \mu \quad .842(.457) = 16.31 - \mu$$

$$-2.124\sigma = -1.97$$

$$\sigma = .457$$

$$\mu = 15.93$$

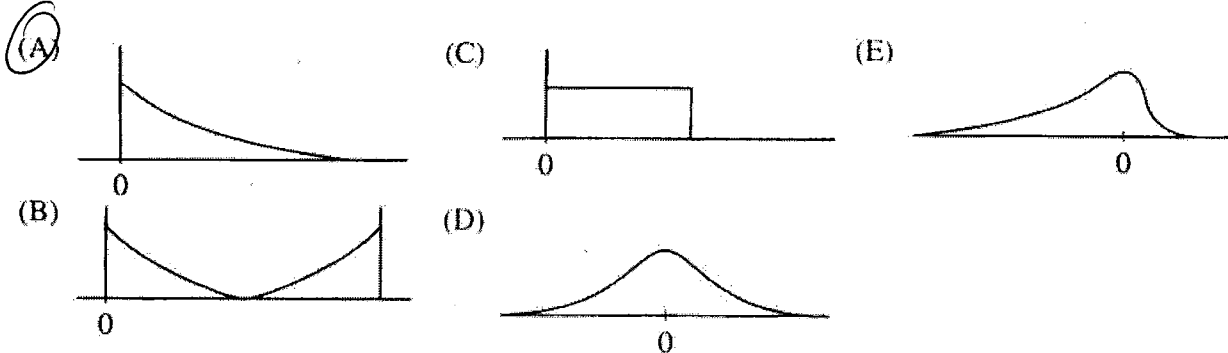
27. The weights of the male and female students in a class are summarized in the following boxplots:



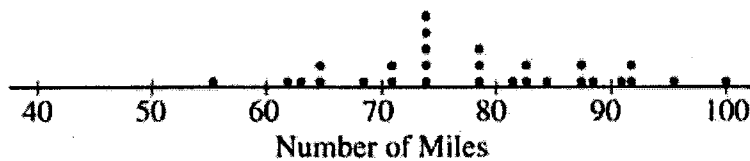
Which of the following is NOT correct?

- (A) About 50% of the male students have weights between 150 and 185 pounds.
- (B) About 25% of female students have weights more than 130 pounds.
- (C) The median weight of male students is about 162 pounds.
- (D) The mean weight of female students is about 120 pounds because of symmetry in the boxplot.
- (E) The male students have less variability than the female students.

28. For which of the following distributions is the mean greater than the median?



29. The dotplot below displays the total number of miles that the 28 residents of one street in a certain community traveled to work in one five-day workweek.



Which of the following is closest to the percentile rank of a resident from this street who traveled 85 miles to work that week?

- (A) 60
  - (B) 70
  - (C) 75
  - (D) 80
  - (E) 85
- $\frac{19}{27} = .704$

30. A distribution of test scores is not symmetric. Which of the following is the best estimate of the z-score of the third quartile?

- (A) 0.67
- (B) 0.75
- (C) 1.00
- (D) 1.41
- (E) This z-score cannot be estimated from the information given.

Use the following information to answer #31-33: The heights of American men aged 18 to 24 are approximately normally distributed with mean 70 inches and standard deviation 2.5 inches.

31. Only about 5% of men aged 18 to 24 have heights outside the interval...

- (A) 65 to 75 in      (B) 67.5 to 72.5 in      (C) 62.5 to 77.5 in      (D) 60 to 80 in      (E) None of these

32. If a young man between the ages of 18 and 24 is found to have a z-score of  $-1.23$ , what is his actual height?

$$-1.23 = \frac{x - 70}{2.5} \quad \boxed{66.925 \text{ in}}$$

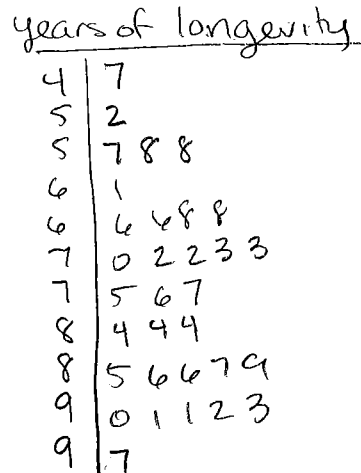
33. If a young man between the ages of 18 and 24 ranks in the 92<sup>nd</sup> percentile for his height, how tall is he?

$$z^*(.92) = 1.405 \quad 1.405 = \frac{x - 70}{2.5} \quad \boxed{73.513 \text{ in}}$$

34. How long did real cowboys live? One answer may be found in the book *The Last Cowboys* by Connie Brooks (University of New Mexico Press). This delightful book presents a thoughtful sociological study of cowboys in West Texas and Southeastern New Mexico around the year 1890. A sample of 32 cowboys gave the following years of longevity:

58    52    68    86    72    66    97    89    84    91    91  
 92    66    68    87    86    73    61    70    75    72    73  
 85    84    90    57    77    76    84    93    58    47

a) Make a split-stem stemplot for these data.



b) Find the mean and the five-number summary of the data.

Mean 75.875    Med 75.5  
 min 47    Q3 86.5  
 Q1 67    Max 97

35. (2011 #1)

A professional sports team evaluates potential players for a certain position based on two main characteristics, speed and strength.

(a) Speed is measured by the time required to run a distance of 40 yards, with smaller times indicating more desirable (faster) speeds. From previous speed data for all players in this position, the times to run 40 yards have a mean of 4.60 seconds and a standard deviation of 0.15 seconds, with a minimum time of 4.40 seconds, as shown in the table below.

	Mean	Standard Deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

Based on the relationship between the mean, standard deviation, and minimum time, is it reasonable to believe that the distribution of 40-yard running times is approximately normal? Explain.

If the distribution was  $\sim N$  the min would be approximately 3 std. dev. from the mean. Since it is between 1 and 2 std. dev, this distribution is not  $\sim N$ .



- (b) Strength is measured by the amount of weight lifted, with more weight indicating more desirable (greater) strength. From previous strength data for all players in this position, the amount of weight lifted has a mean of 310 pounds and a standard deviation of 25 pounds, as shown in the table below.

	Mean	Standard Deviation
Amount of weight lifted	310 pounds	25 pounds

Calculate and interpret the z-score for a player in this position who can lift a weight of 370 pounds.

$$z = \frac{370 - 310}{25} = 2.4$$

This indicates the player can lift 2.4 std. dev above the mean weight lifted for all players in that position.

- (c) The characteristics of speed and strength are considered to be of equal importance to the team in selecting a player for the position. Based on the information about the means and standard deviations of the speed and strength data for all players and the measurements listed in the table below for Players A and B, which player should the team select if the team can only select one of the two players? Justify your answer.

	Player A	Player B
Time to run 40 yards	4.42 seconds	4.57 seconds
Amount of weight lifted	370 pounds	375 pounds

Player A

$$z(\text{lift}) = \frac{370 - 310}{25} = 2.4$$

$$z(\text{run}) = \frac{4.42 - 4.6}{.15} = -1.2$$

Player B

$$z(\text{lift}) = \frac{375 - 310}{25} = 2.6$$

$$z(\text{run}) = \frac{4.57 - 4.6}{.15} = -.2$$

Both players are similar in their lift at 2.4 and 2.6 standard deviations. However, Player A is an entire standard deviation below Player B, making A significantly faster. Therefore, I would choose Player A.

