AP Statistics

Review Week 1

SOCS and GRAPHS

Advanced Placement AAP Review will be held in **room 315** and **312** on Tuesdays and Thursdays.

The week of March 23rd we will be reviewing **SOCS and GRAPHS**.

The session will begin in room 315 with a brief review of the weekly topic.

Instruction will be from 3:15 pm to 3:30 pm

Once we have reviewed the topic you may begin practicing the questions in your review packet.

Answers will be posted in room 315 and 312 all week and will be posted on line after 3:00 pm on Friday the week of review.

If you have difficulty with a question look at the detailed answer postings BEFORE you ask your teacher for help.

Get a hint....**DON'T COPY THE ANSWER!!! THAT IS NOT HELPFUL!!**

When you have completed a question...**REFLECT!!!!** Ask yourself what skill you used to solve that problem and write that down!!

Once we have completed the weekly review, keep it to study from as we get closer to the exam.

SOCS

Brief Review

Describe a distribution using SOCS

SHAPE – Unimodal and Symmetric (Hump in the middle) or skewed left or right (direction of skewness is decided by the tail … catch the tiger by the tail)

Outliers – optional

Center – middle of the hump….Mean or Median

Spread – smallest to largest…Standard Deviation or IQR

Graphs

- Label your axis with an appropriate scale and context.
- Ask yourself if someone could read your graph and understand the problem without reading the problem.

Categorical Data:

- Bar Charts
- Double Bar Charts
- Segmented Bar Charts.
- Circle Graphs – though I have never seen one on an AP exam

Quantitative Data:

- Dotplots
- Stemplots
- Histograms
- Ogives – Cumulative frequency graph
- Boxplots – modified (shows outliers)

Comparing Distributions:

- DON'T USE "*while*"……
- Compare using bigger, smaller, greater, less than…etc.
- CONTEXT CONTEXT CONTEXT!!!!
- Weave the descriptions together.
- Think about what you want to write before you do.
- Identify center and spread as mean, median, standard deviation etc…only if you KNOW that value….include units.

1. The statistics below provide a summary of the distribution of heights, in inches, for a simple random sample of 200 young children.

       Mean: 46 inches
       Median: 45 inches
       Standard Deviation: 3 inches
       First Quartile: 43 inches
       Third Quartile: 48 inches

About 100 children in the sample have heights that are

(A) less than 43 inches
(B) less than 48 inches
(C) between 43 and 48 inches
(D) between 40 and 52 inches
(E) more than 46 inches

---

7. The stemplot below shows the yearly earnings per share of stock for two different companies over a sixteen-year period.

Company A		Company B
	0	58, 75, 96, 98
92, 91, 90, 82, 78, 43, 38, 26	1	01, 10, 17, 21, 43, 43, 53, 65, 73
49, 47, 44, 00	2	09, 27, 29
73, 27, 05, 02	3	

Which of the following statements is true?
(A) The median of the earnings of Company A is less than the median of the earnings of the Company B.
(B) The range of the earnings of Company A is less than the range of the earnings of Company B.
(C) The third quartile of Company A is smaller than the third quartile of Company B.
(D) The mean of the earnings of Company A is greater than the mean of the earnings of Company B.
(E) The interquartile range of Company A is twice the interquartile range of Company B.

---

15. The histograms below represent the distribution of five different data sets, each containing 28 integers, from 1 through 7, inclusive. The horizontal and vertical scales are the same for all graphs. Which graph represents the data set with the largest standard deviation.

(A)

(B)

(C)

(D)

(E)

----
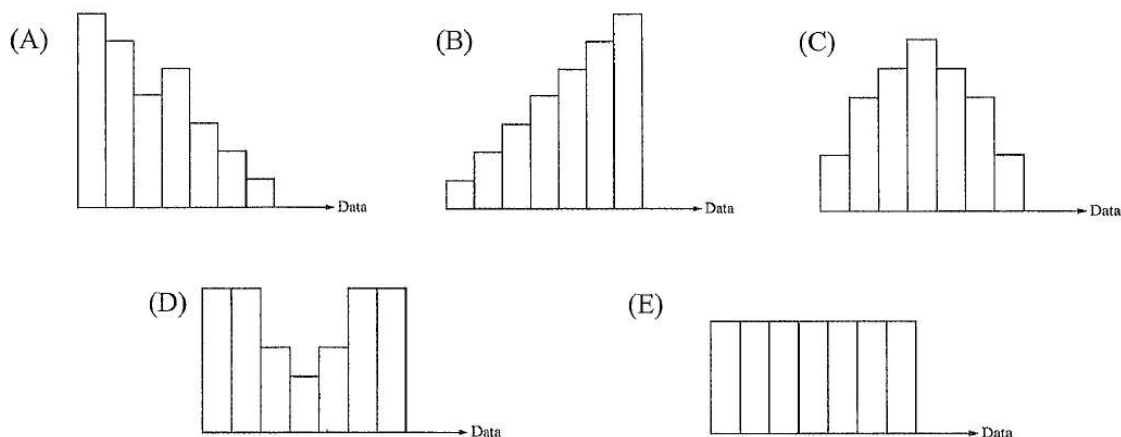
18. One hundred people were interviewed and classified according to their attitude toward small cars and their personality type. The results are shown in the table below.

|		Personality Type		
		Type A	Type B	Total
Attitude	Positive	25	12	37
Toward	Neutral	11	9	20
Small Cars	Negative	24	19	43
	Total	60	40	100

Which of the following is true?

(A) Of the three attitude groups, the group with the negative attitude has the highest proportion of type A personality types.
(B) Of the three attitude groups, the group with the neutral attitude has the highest proportion of type B personality types.
(C) For each personality type, more than half of the 100 respondents have a neutral attitude toward small cars.
(D) The proportion that has a positive attitude toward small cars is higher among people with a type B personality type than among people with a type A personality type.
(E) More than half of the 100 respondents have a type A personality type and a positive attitude toward small cars.

24. The histogram below displays the times, in minutes, needed for each chimpanzee in a sample of 26 to complete a simple navigational task.
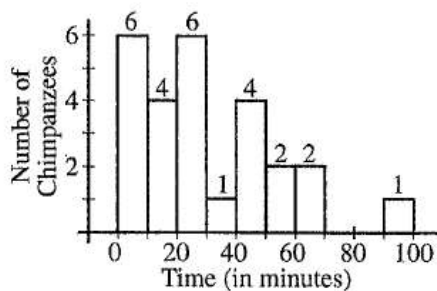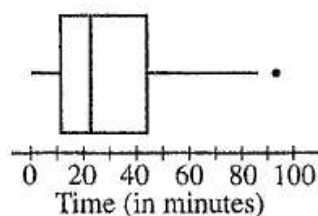

It was determined that the largest observation, 93, is an outlier since $Q_3 + 1.5(Q_3 - Q_1) = 87.125$. Which of the following boxplots could represent the information in the histogram?
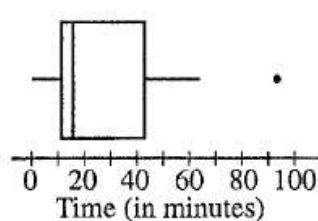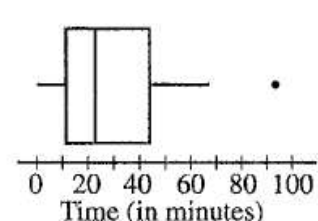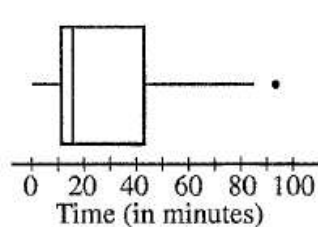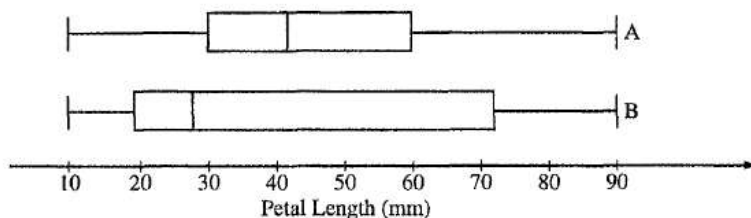
(A)

(B)

(C)

(D)

(E)
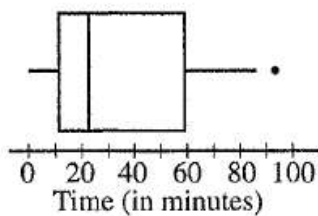
Petal Length (mm)

29. A botanist is studying the petal lengths, measured in millimeters, of two species of lilies. The boxplots above illustrate the distribution of petal lengths from two samples of equal size, one from species A and the other from species B. Based on these boxplots, which of the following is a correct conclusion about the data collected in this study?

(A) The interquartile ranges are the same for both samples.
(B) The range for species B is greater than the range for species A.
(C) There are more petal lengths that are greater than 70 mm for species A than there are for species B.
(D) There are more petal lengths that are greater than 40 mm for species B than there are for species A.
(E) There are more petal lengths that are less than 30 mm for species B than there are for species A.

2003 Exam
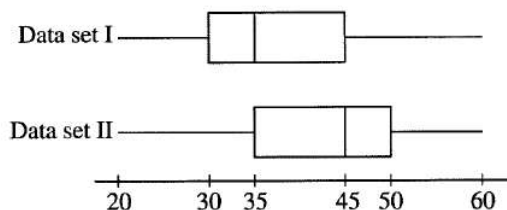
14. The boxplots shown above summarize two data sets, I and II. Based on the boxplots, which of the following statements about these two data sets CANNOT be justified?

(A) The range of data set I is equal to the range of data set II.

(B) The interquartile range of data set I is equal to the interquartile range of data set II.

(C) The median of data set I is less than the median of data set II.

(D) Data set I and data set II have the same number of data points.

(E) About 75% of the values in data set II are greater than or equal to about 50% of the values in data set I.

20. A small town employs 34 salaried, nonunion employees. Each employee receives an annual salary increase of between $500 and $2,000 based on a performance review by the mayor's staff. Some employees are members of the mayor's political party, and the rest are not.

Students at the local high school form two lists, A and B, one for the raises granted to employees who are in the mayor's party, and the other for raises granted to employees who are not. They want to display a graph (or graphs) of the salary increases in the student newspaper that readers can use to judge whether the two groups of employees have been treated in a reasonably equitable manner.

Which of the following displays is <u>least</u> likely to be useful to readers for this purpose?

(A) Back-to-back stemplots of A and B

(B) Scatterplot of B *versus* A

(C) Parallel boxplots of A and B

(D) Histograms of A and B that are drawn to the same scale

(E) Dotplots of A and B that are drawn to the same scale

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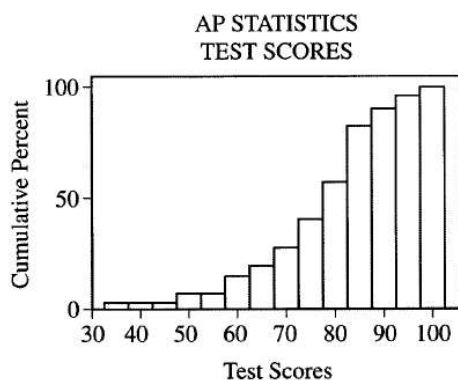
27. The figure above shows a cumulative relative frequency histogram of 40 scores on a test given in an AP Statistics class. Which of the following conclusions can be made from the graph?

(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.

Set A

Set B

10	20	30	40	50
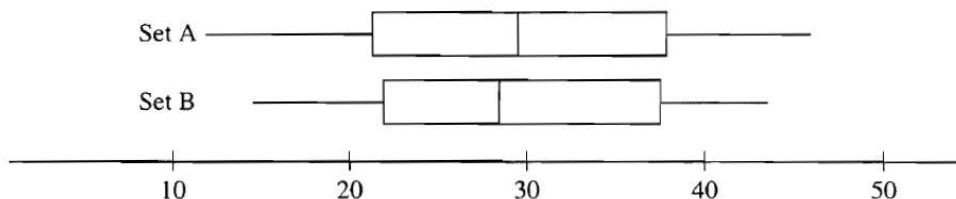
10. The boxplots above summarize two data sets, A and B. Which of the following must be true?

 I. Set A contains more data than Set B.
 II. The box of Set A contains more data than the box of Set B.
 III. The data in Set A have a larger range than the data in Set B.

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

----

21. A company wanted to determine the health care costs of its employees. A sample of 25 employees were interviewed and their medical expenses for the previous year were determined. Later the company discovered that the highest medical expense in the sample was mistakenly recorded as 10 times the actual amount. However, after correcting the error, the corrected amount was still greater than or equal to any other medical expense in the sample. Which of the following sample statistics must have remained the same after the correction was made?

(A) Mean
(B) Median
(C) Mode
(D) Range
(E) Variance

----

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

(A)

(B)

(C)

(D)

(E)
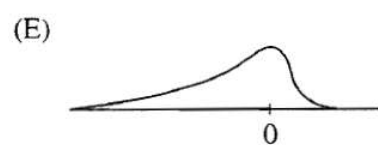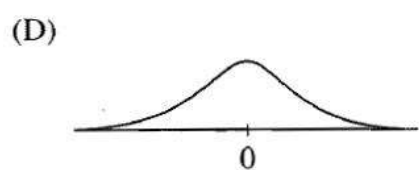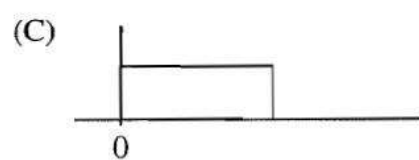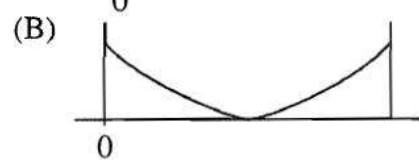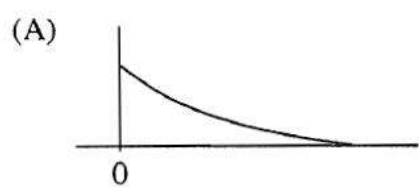
22. The back-to-back stem-and-leaf plot below gives the percentage of students who dropped out of school at each of the 49 high schools in a large metropolitan school district.

```
             School Year                        School Year
              1989-1990                           1992-1993
─────────────────────────────────────      ─────────────────────────────────
                             0 | 0 | 4
                 9 9 9 9 8 8 7 | 0 | 5 6 6 6 7 7 7 8 8 8 9 9
4 4 4 4 4 3 3 2 2 2 2 1 1 1 1 0 | 1 | 0 0 0 0 1 1 1 1 2 2 2 3 3 4 4 4 4
         9 9 9 7 7 6 6 6 6 5 | 1 | 5 5 5 6 6 6 6 7 7 7 7 8
               4 2 2 2 1 0 0 | 2 | 1 3
                   8 8 8 7 6 | 2 |
                           2 | 3 | 0 1 1 2
                       7 6 6 | 3 | 5
                             | 4 |
```
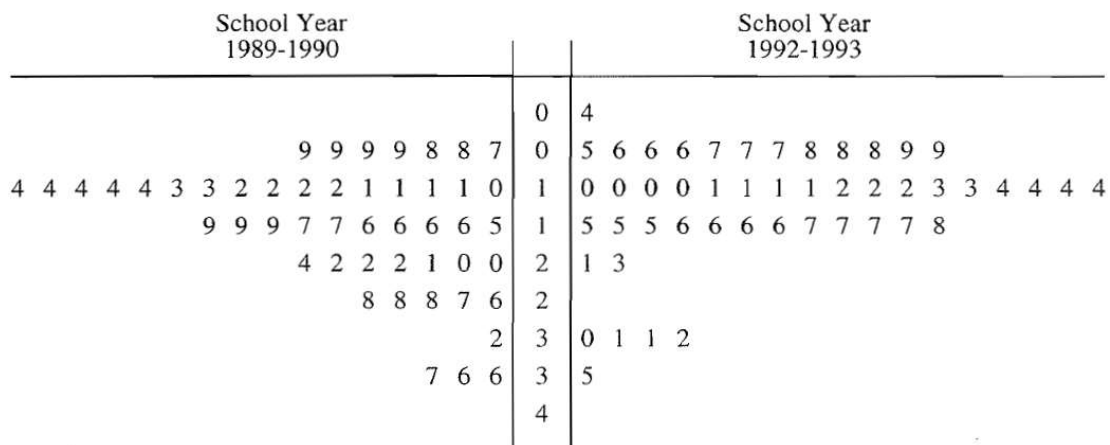
For 1992-1993, 1│2 represents 12%.

FREE RESPONSE PRACTICE

2013 Exam #1

1. An environmental group conducted a study to determine whether crows in a certain region were ingesting food containing unhealthy levels of lead. A biologist classified lead levels greater than 6.0 parts per million (ppm) as unhealthy. The lead levels of a random sample of 23 crows in the region were measured and recorded. The data are shown in the stemplot below.

Lead Levels

```
2 | 8
3 | 0
3 | 5 8 8
4 | 1 1 2
4 | 6 8 8
5 | 0 1 2 2 3 4
5 | 9 9
6 | 3 4
6 | 6 8
```
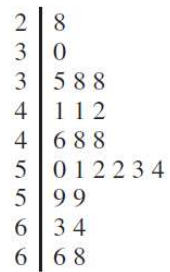
Key: $2|8 = 2.8$ ppm

(a) What proportion of crows in the sample had lead levels that are classified by the biologist as unhealthy?

1. Records are kept by each state in the United States on the number of pupils enrolled in public schools and the number of teachers employed by public schools for each school year. From these records, the ratio of the number of pupils to the number of teachers (P-T ratio) can be calculated for each state. The histograms below show the P-T ratio for every state during the 2001–2002 school year. The histogram on the left displays the ratios for the 24 states that are west of the Mississippi River, and the histogram on the right displays the ratios for the 26 states that are east of the Mississippi River.

West

East

Ratio of the Number of Pupils to the Number of Teachers

Ratio of the Number of Pupils to the Number of Teachers
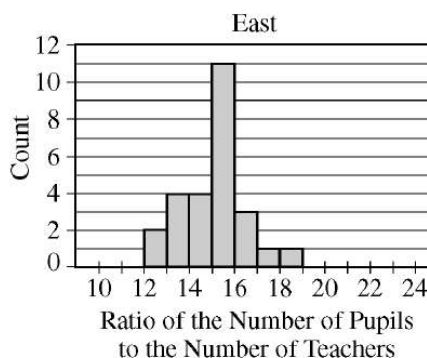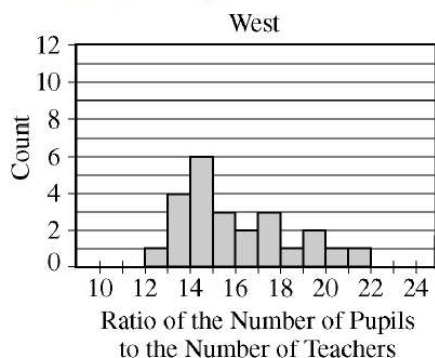
(a) Describe how you would use the histograms to estimate the median P-T ratio for each group (west and east) of states. Then use this procedure to estimate the median of the west group and the median of the east group.

(b) Write a few sentences comparing the distributions of P-T ratios for states in the two groups (west and east) during the 2001–2002 school year.

(c) Using your answers in parts (a) and (b), explain how you think the mean P-T ratio during the 2001–2002 school year will compare for the two groups (west and east).

1. As a part of the United States Department of Agriculture's Super Dump cleanup efforts in the early 1990s, various sites in the country were targeted for cleanup. Three of the targeted sites—River X, River Y, and River Z—had become contaminated with pesticides because they were located near abandoned pesticide dump sites. Measurements of the concentration of aldrin (a commonly used pesticide) were taken at twenty randomly selected locations in each river near the dump sites.

   The boxplots shown below display the five-number summaries for the concentrations, in parts per million (ppm) of aldrin, for the twenty locations that were sampled in each of the three rivers.
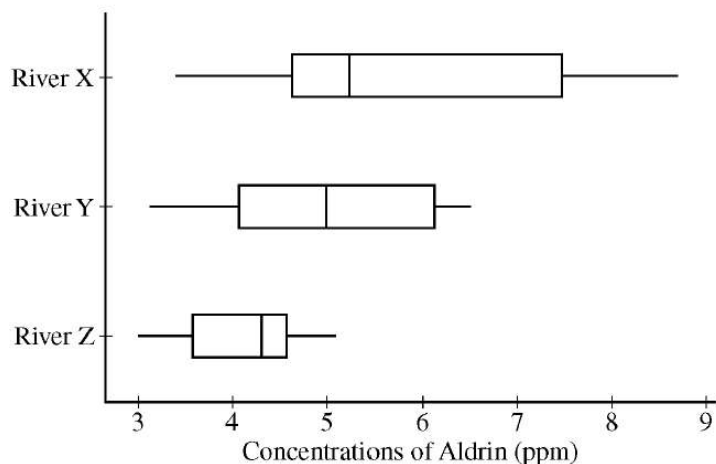
   (a) Compare the distributions of the concentration of aldrin among the three rivers.

   (b) The twenty concentrations of aldrin for River X are given below.

3.4	4.0	5.6	3.7	8.0	5.5	5.3	4.2	4.3	7.3
8.6	5.1	8.7	4.6	7.5	5.3	8.2	4.7	4.8	4.6

   Construct a stemplot that displays the concentrations of aldrin for River X.

   (c) Describe a characteristic of the distribution of aldrin concentrations in River X that can be seen in the stemplot but cannot be seen in the boxplot.

2009 Exam

1. A simple random sample of 100 high school seniors was selected from a large school district. The gender of each student was recorded, and each student was asked the following questions.

1. Have you ever had a part-time job?
2. If you answered yes to the previous question, was your part-time job in the summer only?

The responses are summarized in the table below.

|	Gender		
Job Experience	Male	Female	Total
Never had a part-time job	21	31	52
Had a part-time job during summer only	15	13	28
Had a part-time job but not only during summer	12	8	20
Total	48	52	100

(a) On the grid below, construct a graphical display that represents the association between gender and job experience for the students in the sample.
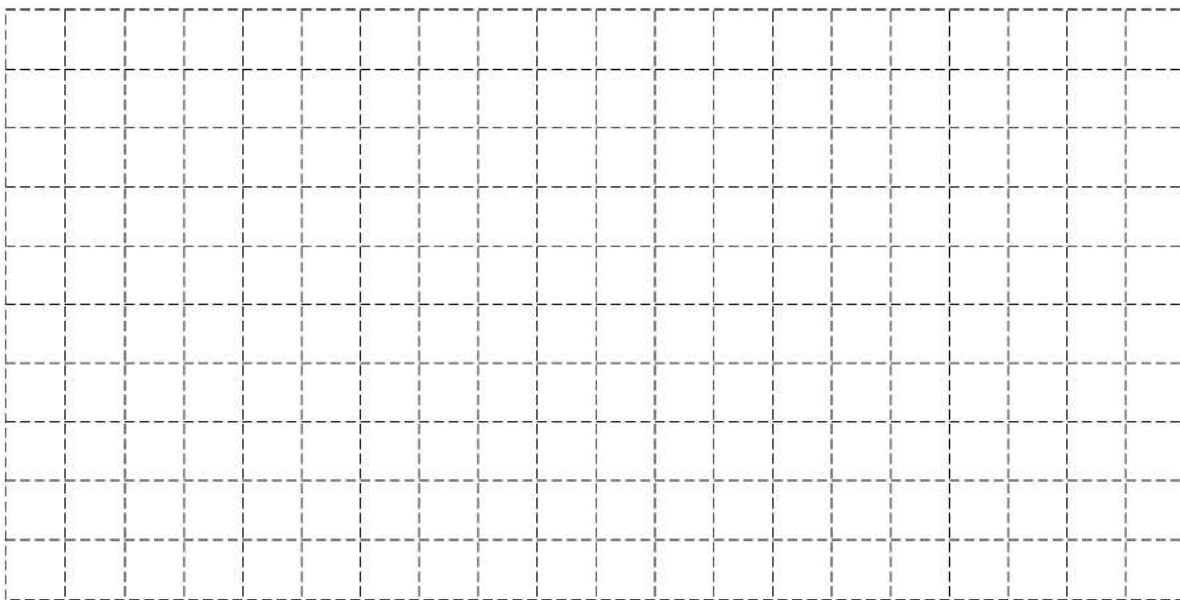
(b) Write a few sentences summarizing what the display in part (a) reveals about the association between gender and job experience for the students in the sample.

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.

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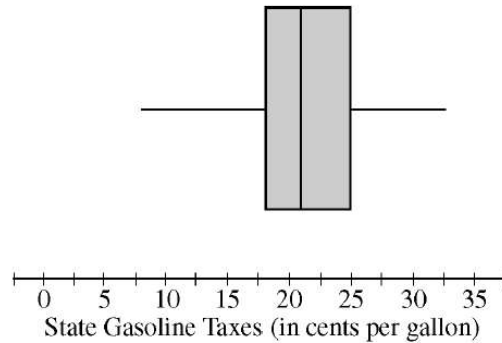
(a) Based on the boxplot, what are the approximate values of the median and the interquartile range of the distribution of state gasoline taxes, in cents per gallon? Mark and label the boxplot to indicate how you found the approximated values.

(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.

1. The department of agriculture at a university was interested in determining whether a preservative was effective in reducing discoloration in frozen strawberries. A sample of 50 ripe strawberries was prepared for freezing. Then the sample was randomly divided into two groups of 25 strawberries each. Each strawberry was placed into a small plastic bag.

   The 25 bags in the control group were sealed. The preservative was added to the 25 bags containing strawberries in the treatment group, and then those bags were sealed. All bags were stored at $0°C$ for a period of 6 months. At the end of this time, after the strawberries were thawed, a technician rated each strawberry's discoloration from 1 to 10, with a low score indicating little discoloration.

   The dotplots below show the distributions of discoloration rating for the control and treatment groups.

   (a) The standard deviation of ratings for the control group is 2.141. Explain how this value summarizes variability in the control group.

   (b) Based on the dotplots, comment on the effectiveness of the preservative in lowering the amount of discoloration in strawberries. (No calculations are necessary.)