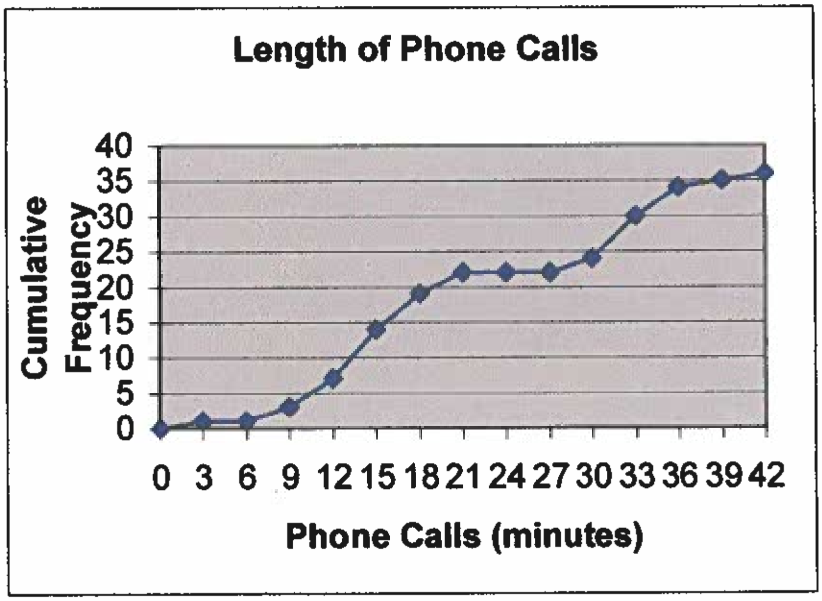
Unp 1 Packet

Name: **AP Statistics**

Special Problem 1A **Gabalot High Phone Calls**

The graph below displays the cumulative frequency of the lengths of phone calls made from the mathematics department office at Gabalot High. Assume that the intervals do not include the number on the left side of the interval. [IE: 0-3 means not 0, but up to and including 3 minute phone calls.]



1. How many phone calls were made from the Gabalot High math office this month? **(1)**

2. Estimate the median length of a phone call. **(1)**

3. What percentage of phone calls lasted more then 30 minutes? **(1)**

4. Construct a boxplot that represents the length of phone calls data. **(2)**

5. Make a frequency table for the length of phone calls data based on the graph above. **(2)**

6. Plot a histogram of these data. **(2)**

7. Describe the distribution of lengths of phone calls made from the math department offices. **(2)**

8. Which plot would you have to adjust to create an ogive? What change would you have to make? **(2)**

Chp 1

Video 2 Guided Notes- The New Normal- Unit 1 Representing Categorical Variables
https://youtu.be/4SXq5H_4Yoo

You may watch the whole 23 min video if you would like.

10:20 min in- starts the context for the Mosaic Plot example
15:20 min in - starts the Mosaic Plot example

10:20 min in
Now let's incorporate a second variable: right vs. left handed. Many people genuinely believe that left-handed people are more artistic than right handed people, but some researchers have tried to prove that false.

Click here to see the elective data sorted by right-handed vs. left-handed. Then copy that table below and answer the questions on the right.

	Right Handed	Left Handed
Art		
Music		
Physical Education		
Foreign Language		
Technology		

% of people who chose **Art**:

% of people who chose **Tech** and are **Right Handed**:

% of **Right Handed** people who chose **Tech**:

Reflect

1. How many variables does this table have? Are they categorical or quantitative?

2. Which variable would best explain or predict the other?

15:20 min in

Side by Side Bar Graph:

Segmented Bar Graph:

Mosaic Plot:

3. How do the bars in the side-by-side-bar graph relate to the bars in the segmented bar graph?

4. How do the bars in the segmented bar graph relate to the bars in the mosaic plot?

5. Is there an association between right vs. left handedness and favorite elective? If so, describe it.

6. If there were not an association between right vs. left handedness and favorite elective, what would the graphs look like? Explain.

Mosaic Plots

For use with the discussion of side-by-side and segmented bar graphs on page 19 (6e).

Yellowstone National Park staff surveyed a random sample of 1526 winter visitors to the park. They asked each person whether he or she belonged to an environmental club (like the Sierra Club). Respondents were also asked whether they owned, rented, or had never used a snowmobile. Here is a two-way table summarizing the results.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

Figure 1.3 compares the distributions of snowmobile use for Yellowstone National Park visitors who are environmental club members and those who are not environmental club members with (a) a **side-by-side bar graph**, (b) a **segmented bar graph**, and (c) a **mosaic plot**. Notice that the segmented bar graph can be obtained by stacking the bars in the side-by-side bar graph for each of the two environmental club membership categories (no and yes). The bar widths in the mosaic plot are proportional to the number of survey respondents who are (305) and are not (1221) environmental club members.

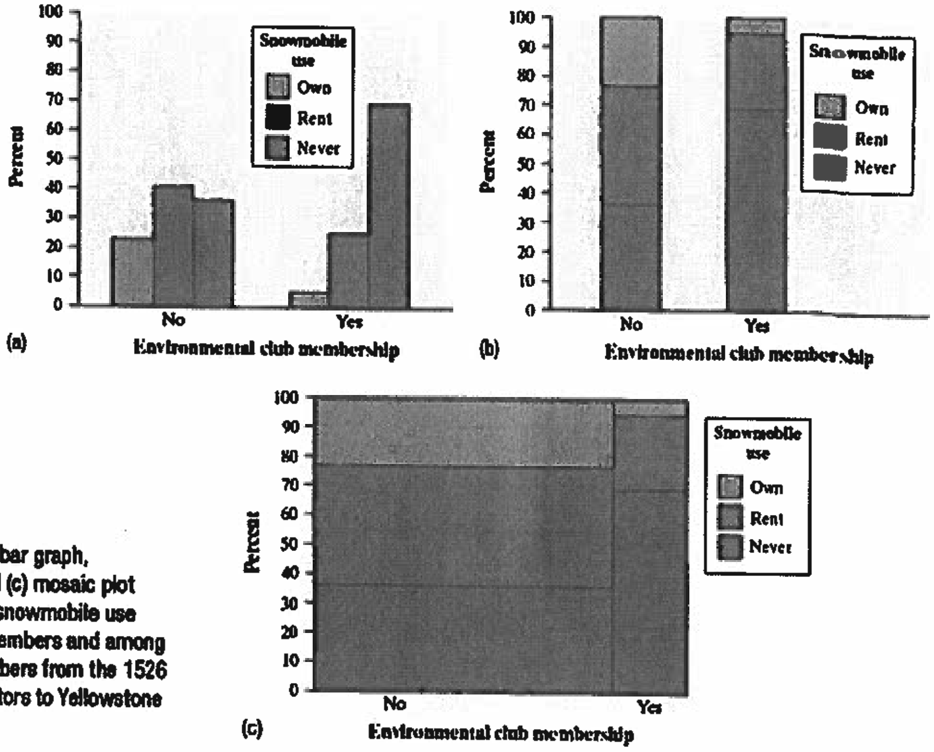


FIGURE 1.3 (a) Side-by-side bar graph, (b) segmented bar graph, and (c) mosaic plot displaying the distribution of snowmobile use among environmental club members and among non-environmental club members from the 1526 randomly selected winter visitors to Yellowstone National Park.

Extracted from Starnes/Tabor, *Updated The Practice of Statistics, 6e*

DEFINITION Side-by side bar graph, Segmented bar graph, Mosaic plot

A **side-by-side bar graph** displays the distribution of a categorical variable for each value of another categorical variable. The bars are grouped together based on the values of one of the categorical variables and placed side by side.

A **segmented bar graph** displays the distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.

A **mosaic plot** is a modified segmented bar graph in which the width of each rectangle is proportional to the number of individuals in the corresponding category.

Extracted from Starnes/Tabor, *Updated The Practice of Statistics, 6e*

Exercises

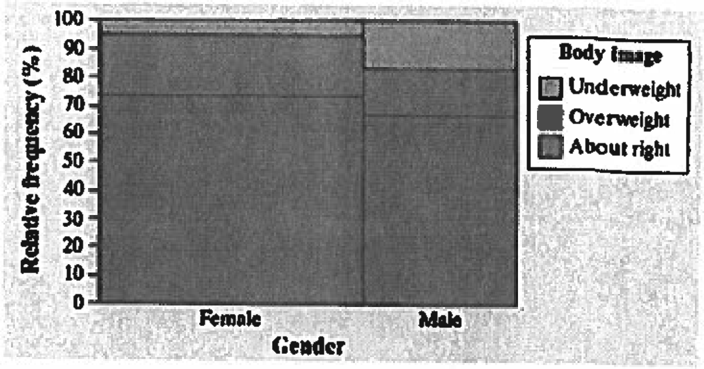
1. **Body image** A random sample of 1200 U.S. college students was asked, "What is your perception of your own body? Do you feel that you are overweight, underweight, or about right?" The two-way table summarizes the data on perceived body image by gender.

		Gender		
		Female	Male	Total
Body image	About right	560	295	855
	Overweight	163	72	235
	Underweight	37	73	110
	Total	760	440	1200

(a) Of the respondents who felt that their body weight was about right, what proportion were female?

(b) Of the female respondents, what percent felt that their body weight was about right?

(c) The mosaic plot displays the distribution of perceived body image by gender. Describe what this graph reveals about the association between these two variables for the 1200 college students in the sample.



Name:

C.1 Review B AP Statistics

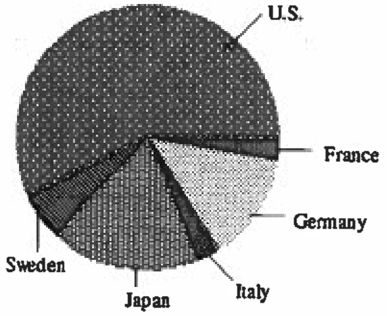
Directions: *Work on these sheets. Answer completely, but be concise.*

Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. Mr. Yates picked up a dozen items in the grocery store with a mean cost of $3.25. Then he added an apple pie for $6.50. The new mean for all 13 items is

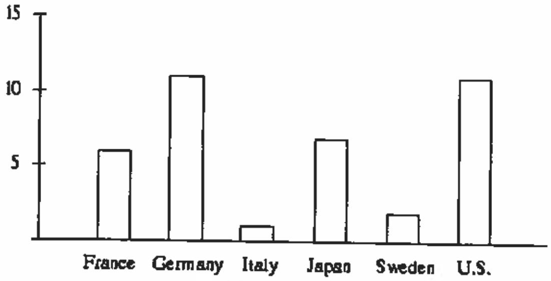
(a) $3.00 (b) $3.50 (c) $3.75 (d) $4.88 (e) None of the above

Use the following to answer Question 2:

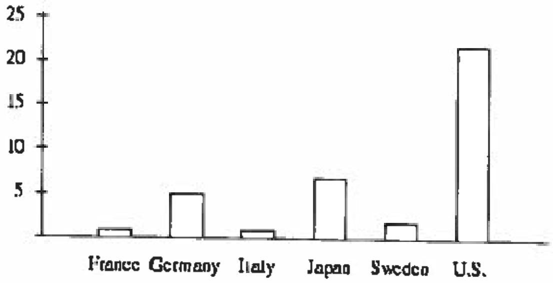


2. Which of the following bar graphs is equivalent to the pie chart?

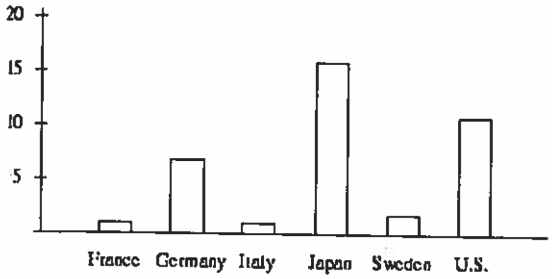
(a)



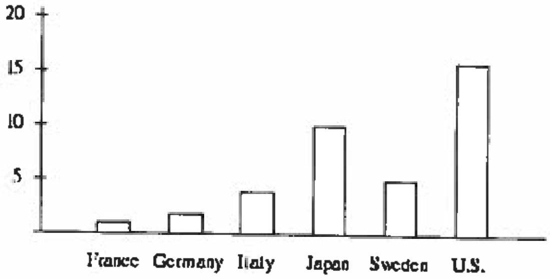
(b)



(c)

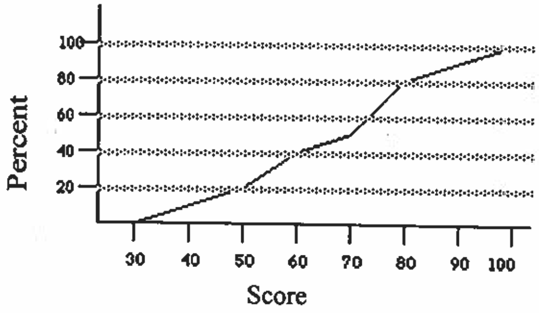


(d)



(e) None of these.

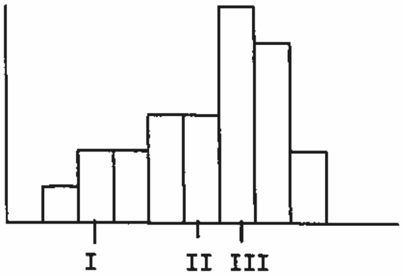
3. Consider the following ogive of the scores of students in an introductory statistics course:



A grade of C or C+ is assigned to a student who scores between 55 and 70. The percentage of students who obtained a grade of C or C+ is

(a) 25% (b) 30% (c) 20% (d) 50% (e) 15%

4. For the following histogram, what is the proper ordering of the mean and median? Note that the graph is NOT numerically precise—only the relative positions are important.



(a) I is the mean and II is the median.
(b) II is the median and III is the mean.
(c) I is the median and II is the mean.
(d) II is the mean and III is the median.
(e) I is the mean and III is the median.

5. A researcher wishes to calculate the average height of patients suffering from a particular disease. From patient records, the mean was computed as 156 cm, and standard deviation as 5 cm. Further investigation reveals that the scale was misaligned, and that all readings are 2 cm too large, for example, a patient whose height is really 180 cm was measured as 182 cm. Furthermore, the researcher would like to work with statistics based on meters. The correct mean and standard deviation ar:
(a) 1.56m, 0.05m
(b) 1.54m, 0.05m
(c) 1.56m, 0.03m
(d) 1.58m, 0.05m
(e) 1.58m, 0.07m

6. A medical researcher collects health data on many women in each of several countries. One of the variables measured for each woman in the study is her weight in pounds. The following list gives the five-number summary for the weights of women in one of the countries.

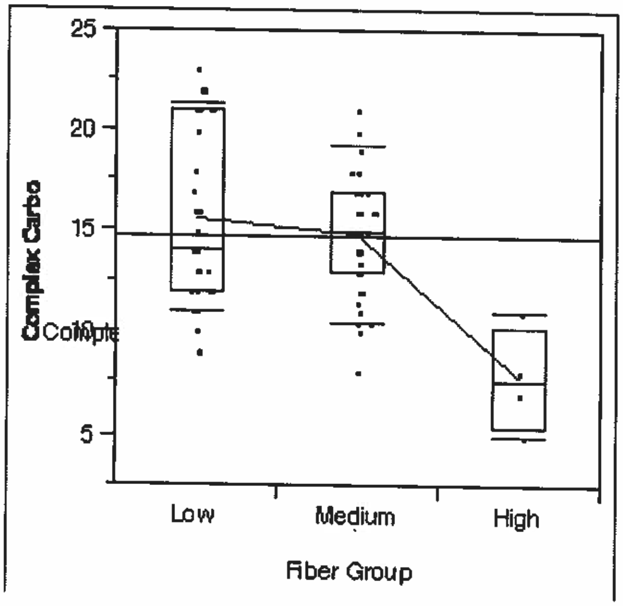
Country A: 100, 110, 120, 160, 200

About what percent of Country A women weigh between 110 and 200 pounds?
(a) 50%
(b) 65%
(c) 75%
(d) 85%
(e) 95%

7. The median age of five people in a meeting is 30 years. One of the people, whose age is 50 years, leaves the room. The median age of the remaining four people in the room is

(a) 40 years.
(b) 30 years.
(c) 25 years.
(d) less than 30 years.
(e) Cannot be determined from the information given.

9. Here is a summary graph of complex carbohydrates (in grams) for each of three fiber groups in a set of data related to cereals.



Which of the following is NOT correct?

(a) The low-fiber group is more variable than the medium-fiber group because the central box is larger.

(b) About 25% of low-fiber cereals have less than 12 g of complex carbohydrates per serving.

(c) About 50% of medium-fiber cereals have more than 15 g of complex carbohydrates per serving.

(d) The average amount of complex carbohydrates per serving for the high-fiber group appears to be much smaller than for the other two groups.

(e) About 25% of the medium-fiber cereals have less than 10 g of complex carbohydrates.

10. Earthquake intensities are measured using a device called a seismograph, which is designed to be most sensitive to earthquakes with intensities between 4.0 and 9.0 on the open-ended Richter scale. Measurements of nine earthquakes gave the following readings:

4.5 L 5.5 H 8.7 8.9 6.0 H 5.2

where L indicates that the earthquake had an intensity below 4.0 and H indicates that the earthquake had an intensity above 9.0. The median earthquake intensity of the sample is

(a) Cannot be computed since all of the values are not known

(b) 8.70

(c) 5.75

(d) 6.00

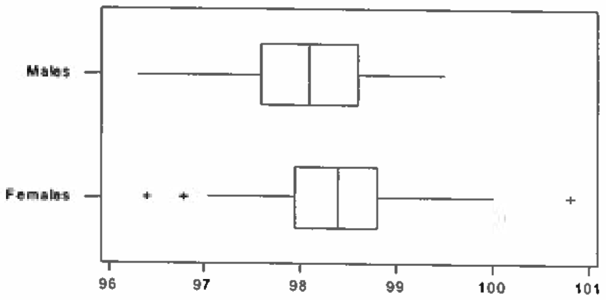
(e) 6.47

Part 2: Free Response

Communicate your thinking clearly and completely.

11. We all "know" that the body temperature of a healthy person is 98.6°F. In reality, the actual body temperature of individuals varies. Here is a back-to-back stemplot of the body temperatures of 130 healthy individuals (65 males and 65 females).

(a) Here are boxplots, produced by Minitab, for these distributions. Label both boxplots with the five-number summary values.



Males		Females
3	96	
	96	4
7	96	7
9	96	8
1110	97	
32	97	22
544444	97	4
7666	97	677
998888	97	8888999
11000000	98	000001
332222	98	222222333
554444	98	444445
77666666	98	6666777777
9888	98	8888889
1000	99	0011
32	99	223
54	99	4
	99	
	99	9
	100	0
	100	
	100	
	100	
	100	8

(b) Determine whether the 3 points graphed by the + symbol are indeed outliers by our defined criteria.

(c) Write a few sentences comparing the body temperatures of adult males and females.

12. The following data represent scores of 50 students on a calculus test.

72	72	93	70	59	78	74	65	73	80
57	67	72	57	83	76	74	56	68	67
74	76	79	72	61	72	73	76	67	49
71	53	67	65	100	83	69	61	72	68
65	51	75	68	75	66	77	61	64	74

(a) Construct a *relative frequency* histogram for this data set.

(b) Describe the shape, center, and spread of the distribution of test scores.

(c) Would the mean and standard deviation be appropriate measures of center and spread for these test scores? Explain.

Name:

C.1 Review C AP Statistics

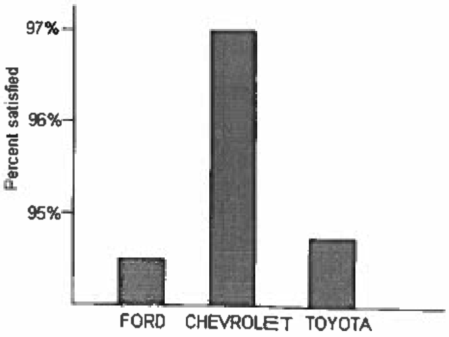
Directions: *Work on these sheets. Answer completely, but be concise.*

Part I: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380 students took the test. About how many had scores between 35 and 61?

(a) 26 (b) 76 (c) 95 (d) 190 (e) None of these

2. The following bar graph gives the percent of owners of three brands of trucks who are satisfied with their truck.



From this graph we may legitimately conclude that

(a) owners of other brands of trucks are less satisfied than the owners of these three brands.
(b) Chevrolet owners are substantially more satisfied than Ford or Toyota owners.
(c) there is very little difference in the satisfaction of owners for the three brands.
(d) Chevrolet probably sells more trucks than Ford or Toyota.
(e) a pie chart would have been a better choice for displaying these data.

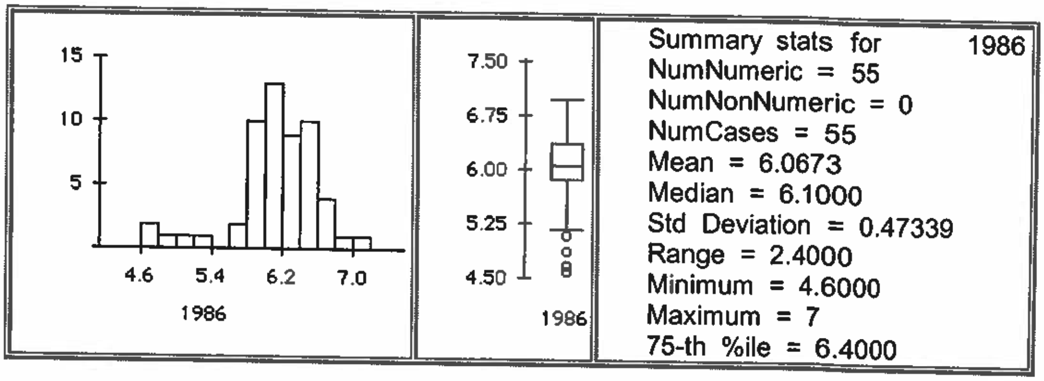
3. A reporter wishes to portray baseball players as overpaid. Which measure of center should he report as the average salary of major league players?

(a) The mean.
(b) The median.
(c) Either the mean or median. It doesn't matter since they will be equal.
(d) Neither the mean nor median. Both will be much lower than the actual average salary.
(e) The standard deviation should be used to show the great disparity between the astronomical salaries of the few superstars and the salaries of the rest of the players.

4. The mean salary of all female workers is $35,000. The mean salary of all male workers is $41,000. What must be true about the mean salary of all workers?

(a) It must be $38,000.
(b) It must be larger than the median salary.
(c) It could be any number between $35,000 and $41,000.
(d) It must be larger than $38,000.
(e) It cannot be larger than $40,000.

5. Consider the following output analyzing pH values of some 1986 data on precipitation events.



```
Summary stats for          1986
NumNumeric = 55
NumNonNumeric = 0
NumCases = 55
Mean = 6.0673
Median = 6.1000
Std Deviation = 0.47339
Range = 2.4000
Minimum = 4.6000
Maximum = 7
75-th %ile = 6.4000
```

Which of the following is NOT correct?

(a) The 25th percentile is about 5.9.
(b) Some outliers appear to be present below a pH of 5.2.
(c) About 95% of the observations have pH values in the approximate range 6 ± 1.
(d) About 10% of the values are in the range 5.8 to 6.0.
(e) About 75% of the values are less than 6.4.

6. A sample of 99 distances has a mean of 24 feet and a median of 24.5 feet. Unfortunately, it has just been discovered that an observation which was erroneously recorded as "30" actually had a value of "35." If we make this correction to the data, then
(a) the mean remains the same, but the median is increased.
(b) the mean and median remain the same.
(c) the median remains the same, but the mean is increased.
(d) the mean and median are both increased.
(e) we do not know how the mean and median are affected without further calculations, but the variance is increased.

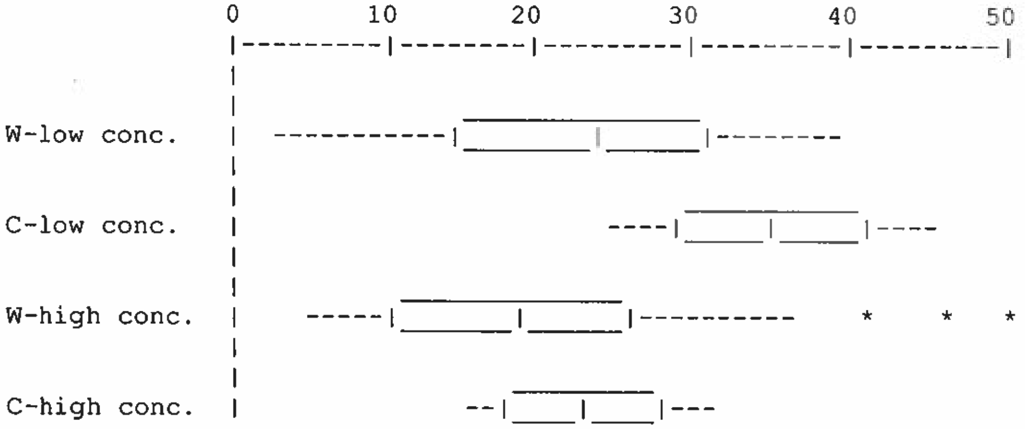
7. Forty students took a statistics examination having a maximum of 50 points. The score distribution is given in the following stem-and-leaf plot:

```
0|28
1|2245
2|01333358889
3|001356679
4|22444466788
5|000
```

The third quartile of the score distribution is equal to
(a) 43 (b) 44 (c) 45 (d) 23 (e) 32

8. Rainwater was collected in water collectors at 30 different sites near an industrial comples and the amount of acidity (pH level) was measured. The mean and standard deviation of the values are 4.60 and 1.10, respectively. When the pH meter was recalibrated back at the laboratory, it was found to be in error. The error can be corrected by adding 0.1 pH units to all of the values and then multiplying the result by 1.2. The mean and standard deviation of the corrected pH measurements are
(a) 5.64, 1.44 (b) 5.64, 1.32 (c) 5.40, 1.44 (d) 5.40, 1.32 (e) 5.64, 1.20

9. An experiment was conducted to investigate the effect of a new weed killer to suppress weed germination in onion crops. Two chemicals were used, the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations. Measurements are made, of the percent weed germination on each of 50 plots for each treatment combination. Here are some boxplots of the results:



Which of the following is NOT a feature of these data?
(a) At either high or low concentrations, the new chemical (W) gives better control of weed germination than the standard weed killer (C).
(b) Fewer weeds germinate at higher concentrations of both chemicals.
(c) The results from the standard chemical are less variable than those from the new chemical.
(d) High or low concentrations of either chemical have approximately the same effects on weed germination.
(e) Some of the results from the low concentration of weed killer W have fewer weeds germinating than some of the results from the high concentration of W.

10. A clothing and textiles student is trying to assess the effect of a jacket's design on the time it takes preschool children to put the jacket on. In a pretest, she times 7 children as they put on her prototype jacket. The times (in seconds) are provided below.

n n 65 39 n 43 102

The n's represent children who had not put the jacket on after 120 seconds (in which case the children were allowed to stop). Which of the following would be the best value to use as the "typical" times required to put on the jacket?

(a) The mean time, which was 62.25 seconds.

(b) The mean time, which was 85.6 seconds.

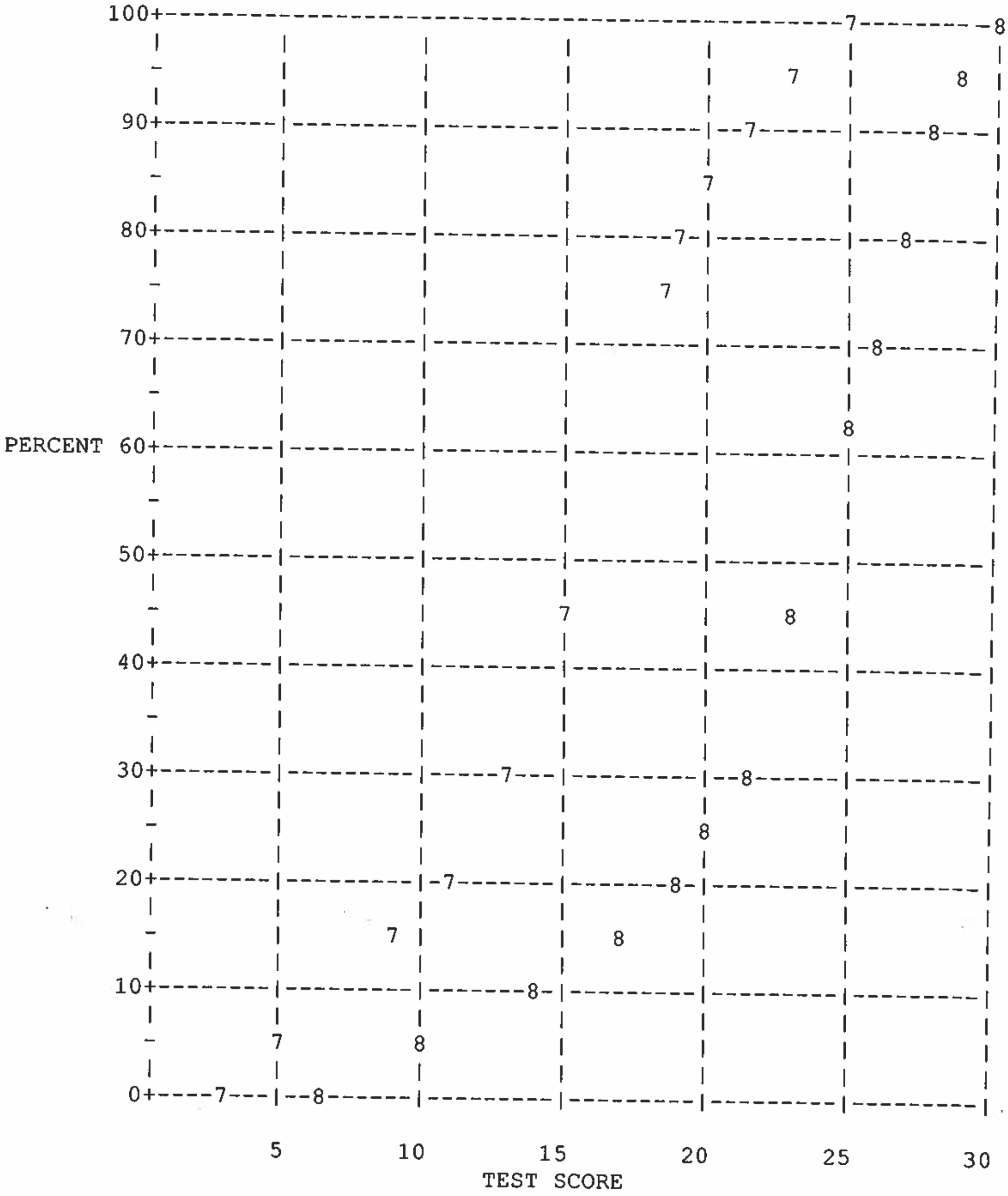
(c) The median time, which was 54 seconds.

(d) The median time, which was 102 seconds.

(e) The missing times (the n's) mean we can't calculate any useful measures of center.

Part 2: Free Response

11. Ogives of Distributions of Arithmetic Test Scores for Seventh- and Eighth-Graders
(Connect 7s with a smooth curve for seventh-grade ogive and connect 8s with a smooth curve for eighth-grade ogive.)



14

(a) What is the estimated percent of eighth-grade pupils whose arithmetic scores fall below the median score for grade seven? Justify your answer.

(b) What is the shape of the distribution of the eighth-grade test scores? Justify your answer.

12. During the early part of the 1994 baseball season, many sports fans and baseball players noticed that the number of home runs being hit seemed to be unusually large. Here are the data on the number of home runs hit by American and National League teams:

American League	35, 40, 43, 49, 51, 54, 57, 58, 58, 64, 68, 68, 75, 77
National League	29, 31, 42, 46, 47, 48, 48, 53, 55, 55, 55, 63, 63, 67

(a) Construct an appropriate graph for comparing the number of home runs hit in the two leagues.

(b) Calculate numerical summaries of the number of home runs hit in the two leagues. Which of these numbers would be most appropriate for comparing the two leagues? Explain.

(c) Are there any outliers in either of the two data sets? Justify your answer numerically.

(d) Write a few sentences comparing the distributions of home runs in the two leagues.