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

Part I: Multiple Choice. *Circle the letter corresponding to the best answer.*
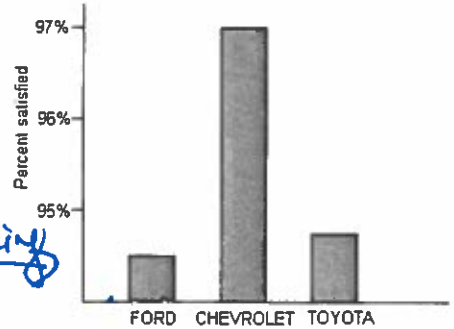
1. The five-number summary for scores on a statistics exam is 11, 35, 61, 70, 79. In all, 380 students took the test. About how many had scores between 35 and 61?

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

   $$\begin{array}{r} 380 \\ \times .25 \\ \hline 95 \end{array}$$

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.

   misleading

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.

   High outliers pull the mean in their direction.

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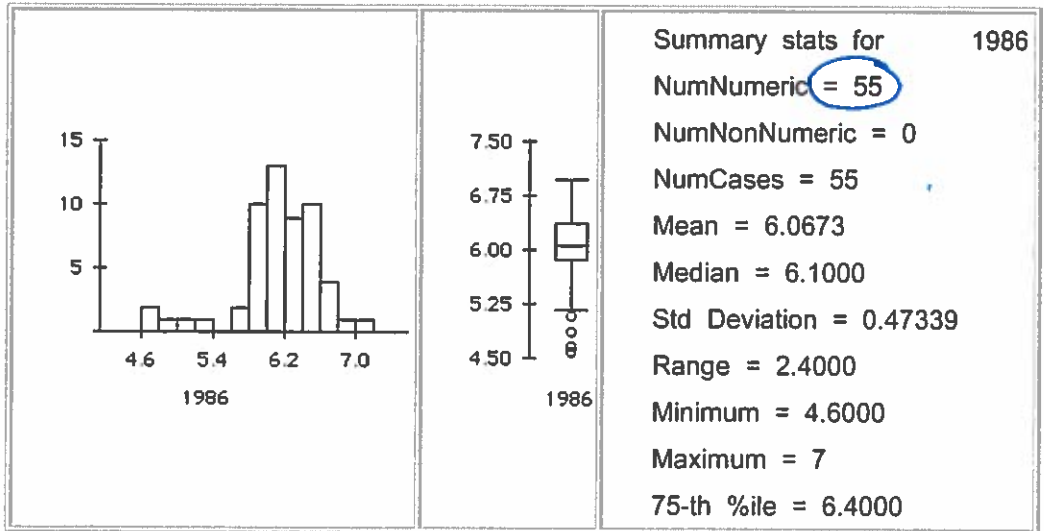   (a) It must be $38,000.
   (b) It must be larger than the median salary.
   (c) It could be any number between $35,000 and $41,000.
   (d) It must be larger than $38,000.
   (e) It cannot be larger than $40,000.

   We don't know how many females + males there are.

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

   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.

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	

   $.10(55) = 5.5$

   $5.8 \to 6$ is 10 pieces of data.

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.

   [handwritten: median is resistant since 30+35 are BOTH above the median, no change]

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:

   [handwritten: n = 40, max 50, 20 20, 1010, 10 10]

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

   [handwritten: mean (4.6 + .1) 1.2 = 5.64, st. dev. 1.1 (1.2) = 1.32]

   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:

   ```
            0         10        20        30        40        50
            |---------|---------|---------|---------|---------|
            |
            |
   W-low conc.  |    -----------|_____|____|---------
            |
   C-low conc.  |                  ----|____|___|----
            |
   W-high conc. |    ----|_____|___|_____|---------    *   *   *
            |
   C-high conc. |            --|___|__|___|---
   ```

   [handwritten: new, old, new, old]
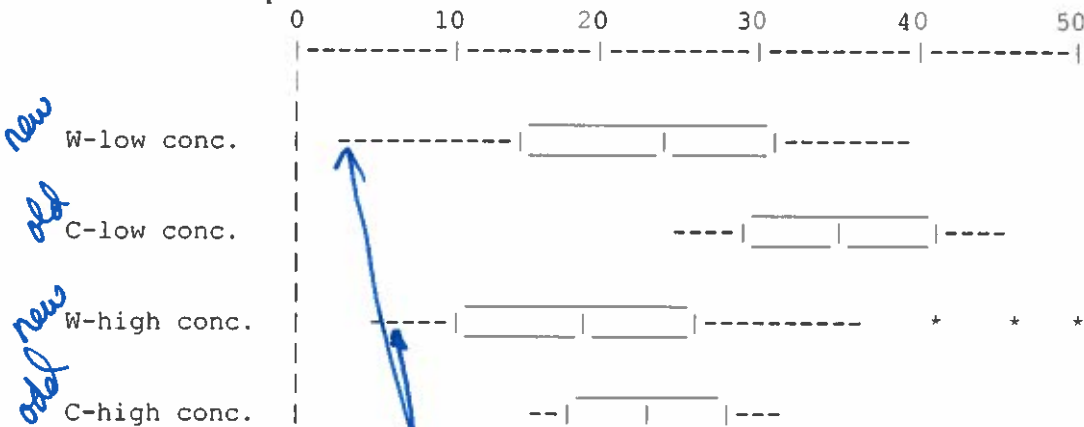
   Which of the following is NOT a feature of these data?
   (a) At either high or low concentrations, the new chemical (W) gives better control of weed germination than the standard weed killer (C). *TRUE*
   (b) Fewer weeds germinate at higher concentrations of both chemicals. *TRUE*
   (c) The results from the standard chemical are less variable than those from the new chemical. *TRUE*
   (d) High or low concentrations of either chemical have approximately the same effects on weed germination. *NO!!*
   (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. *TRUE*

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.

[handwritten: 39, 43, 65 (102), n, n, n]

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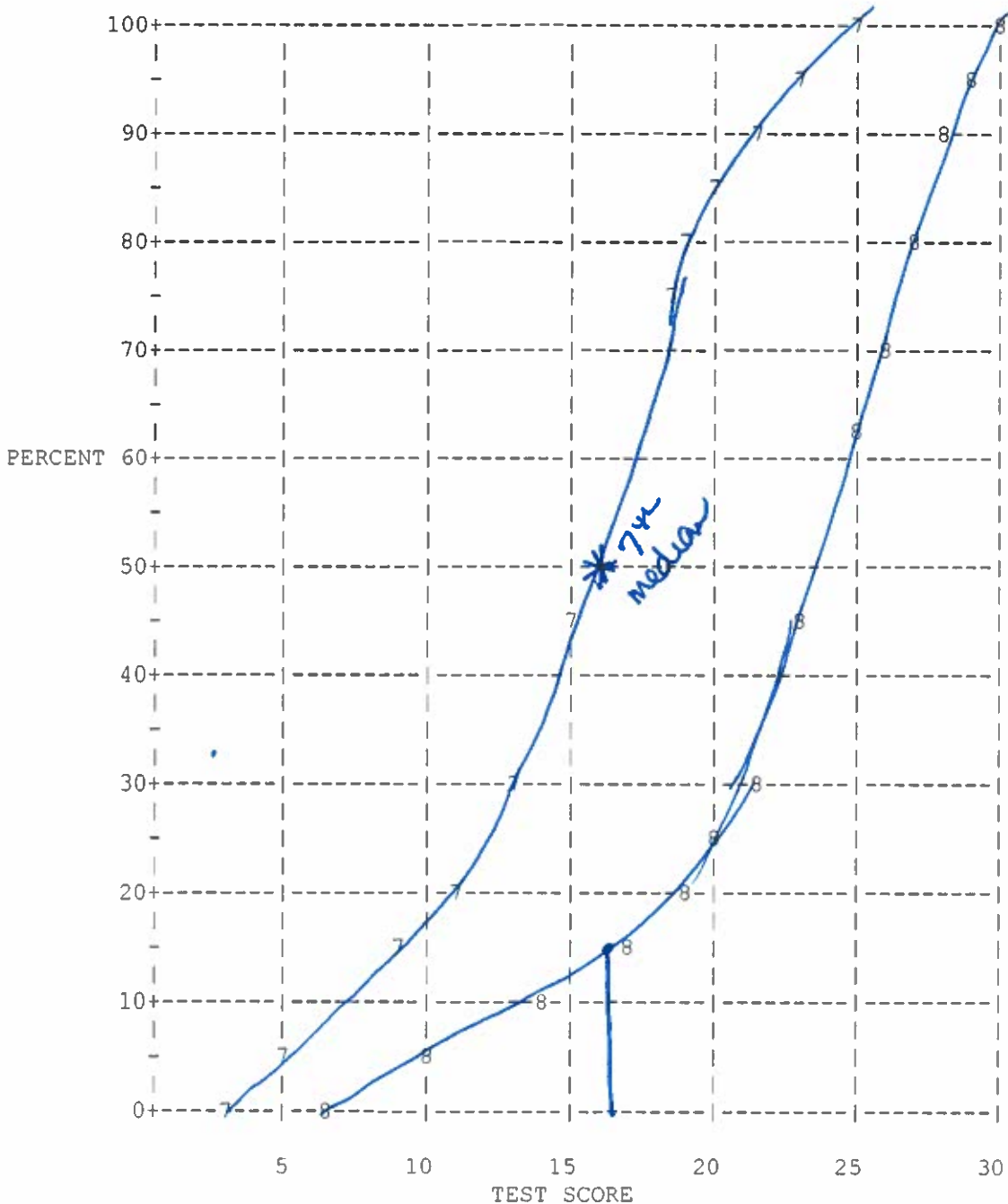
(a) What is the estimated percent of eighth-grade pupils whose arithmetic scores fall below the median score for grade seven? Justify your answer.

about 15% of the 8th graders have scores below 16, which is the median 7th grade score.

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

Fewer of the students have the lower scores. These rapid increase after a score of 20 implies a left skew.

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.

American League   35   49   57.5   68   77

National League 29 31   42   46 50.5 55   67

1994 Home Runs
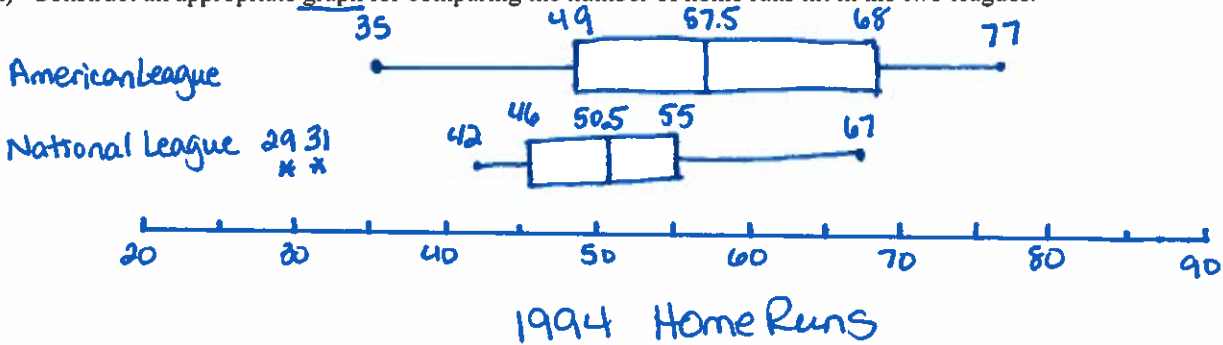
(b) Calculate numerical summaries of the number of home runs hit in the two leagues. Which of these numbers would be most appropriate for comparing the two leagues? Explain.

	min	Q1	med.	Q3	max	$\bar{x}$	s	IQR	Range
AL	35	49	57.5	68	77	56.9	12.7	19	42
NL	29	46	50.5	55	67	50.1	11.1	9	38

Even though there are outliers for the NL, both sets of data are fairly symmetric so we can use the mean to compare the data and st. dev. to compare spread.

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

AL
IQR 19
1.5 IQR 28.5
49 - 28.5 = 20.5   68 + 28.5 = 96.5

NL
9
13.5
46 - 13.5 = 32.5
55 + 13.5 = 68.5

• *No outliers for the AL.*
• *NL has 2 outliers, 29 & 31 (both are below 32.5.)*

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

Both distributions are fairly symmetric with outliers in the National League. The center for the AL (57.5) is higher than the NL (50.5 Home runs).

The ranges (spreads) are about the same unless you ignore the outliers. Then the NL has a smaller range AND st. dev.