

## **Part 1**

- 1. e** Age and number of years at current address are quantitative, gender, favorite subject, and college intent are categorical.
- 2. d** We tend to make comparisons based on area even when height is the intended comparator.
- 3. e** Since there are now only four people in the room, the median is the average of the second and third highest ages. We know that the third oldest is 30, but without knowing the second oldest, we can't determine the median.
- 4. e** "Fairly-priced" is about 40% of the distribution and the percentage for "Don't know" is very low. This reduces the choices to (a) and (e). Percentage for "Expensive" is slightly higher than for "Low cost," which makes (e) correct.
- 5. c** That the mean weight loss is 10.4 does not preclude the possibility of values for weight loss that are zero or even negative (representing a weight gain).
- 6. c** For a group of 50 scores, the third quartile will be the 13<sup>th</sup> highest score, which is in the 30-40 class.
- 7. d** Since the distribution is skewed right, the resistant measures, median and interquartile range, are appropriate.
- 8. d** Standard deviation is a type of average of how much each month's change in unemployment differs from the mean.
- 9. b**  $Q_3$  for males is equal (or nearly equal) to the maximum for females.
- 10. e** Whatever the "bdl" measurements are, they are smaller than 5 ppm, so the median is 7 ppm. We can't know the mean because we don't know the actual values of the "bdl" measurements.

## Part 2

11. (a) The distribution of damages ranges from near \$0 to near \$90 million and is strongly skewed to the right. This suggests that many states suffered very little damage, but a small number of states suffered a great deal of damage. (b) Min. = 0 Q1 = 2.23 Med. = 12.66

Q3 = 41.63 Max. = 88.6. The distance from the median to Q3 is much greater than the distance from the median to Q1, and the distance from Q3 to the maximum is much greater than the distance from Q1 to the minimum. Both of these comparisons suggest that the distribution is skewed right. (c)  $IQR = 41.63 - 2.23 = 39.40$ ;  $1.5 \times IQR = 59.10$ .  $Q1 - 59.10 < 0$  so there are no low outliers, and  $Q3 + 59.10 = 100.73$ , so there are no high outliers. 12. (a) Stemplot at right. (b) There are high values at 195 and above. These scores will cause the mean to be higher, but will have no effect on the median. (c) Mean and standard deviation will increase, median and interquartile range will not change. For anyone watching the movie, life will imitate art.

13. Answers will vary. One effective method for presenting data would be conditional distributions based on gender (row percentages--given in the table below). Any response should discuss the moderate association between gender and means of transportation: a slightly higher proportion of females car pool. 14. Distribution B has the largest standard deviation, since standard deviation measures how far data values are, on average, from the mean, and a greater proportion of the values in B are far from the mean.

	Drive alone	Car Pool	Public trans.
Male	83.7%	11.9%	4.4%
Female	73.9%	20%	6.1%

9	6	Key: 9 6 = 96 min.
10	2 6 9	
11	2 5 8	
12	1 4	
13	0 3 6 9	
14	2 6	
15	2 4	
16	1 5	
17	5 8	
18		
19	5	
20	0 1 7	