1. A zoologist studying adult bears measures a number of different variables. Which of the following possible variables is categorical?

A. the weight in pounds of an adult bear

AR. Incorrect. This is a quantitative variable. The zoologist is measuring a quantity (weight) on a scale of units.

\*B. the level of aggression (low, moderate, high) displayed by an adult bear

BR. Correct. The zoologist is categorizing each unit (bear) according to the three categories of level of aggression, as opposed to counting or measuring a quantity.

C. The number of fish an adult bear eats in a particular day.

CR. Incorrect. This is a quantitative variable—a count of the number of fish.

2. A zoologist studying adult bears measures a number of different variables. One variable is the body temperature of the bear during hibernation. Which of the following is the best description of the distribution of this variable.

\*A. All the values that the zoologist records for body temperature and how many individual bears have each value.

AR. Correct. The distribution of a variable tells us what values a variable takes and how often it takes these values.

B. The difference between the highest temperature recorded for a bear’s body temperature and the lowest.

BR. Incorrect. This difference is called the variable’s range.

C. The geographical area in which the adult bears can be found.

CR. Incorrect. This is what the zoologist would call the distribution of the bear species itself, but it’s not the distribution of the variable “body temperature.”

3. Suppose we classify 315 randomly selected college students according to their general major field and their self-described political viewpoint. The table presents the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Sciences** | **Business** | **Humanities** | **Social Sciences** |
| **Liberal** | 17 | 12 | 32 | 30 |
| **Moderate** | 33 | 40 | 23 | 20 |
| **Conservative** | 35 | 38 | 17 | 18 |

Which of the following list of numbers is a marginal distribution of the variable political viewpoint?

A. 30, 20, 18

AR. Incorrect. This the the distribution of political viewpoint for social science majors only. It’s not a marginal distribution.

B. 85, 90, 72, 68

BR. Incorrect. This is the marginal distribution of hte variable general major field.

\*C. 91, 116, 108

CR. Correct. This is the distribution of political viewpoint with out regard to general major field. For example, 91 is the sum of 17 + 12 + 32 + 30.

4. Suppose we classify 315 randomly selected college students according to their general major field and their self-described political viewpoint. The table presents the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Sciences** | **Business** | **Humanities** | **Social Sciences** |
| **Liberal** | 17 | 12 | 32 | 30 |
| **Moderate** | 33 | 40 | 23 | 20 |
| **Conservative** | 35 | 38 | 17 | 18 |

What percentage of all students surveyed are conservatives majoring in business?

\*A. 12.1%

AR. Correct. We desire the cell percentage for the cell “conservative and moderate.” There are 38 individuals in this cell, so the correct percentage is 38/315 = 0.121, or 12.1%.

B. 42.2%

BR. Incorrect. You have calculated the conditional distribution of viewpoint for the category “conservative” among students majoring in business. We want the percentage for all students, not just students majoring in business.

C. 28.6%

CR. Incorrect. You have calculated the marginal distribution of viewpoint for the category “conservative.” We are interested in onlyconservatives majoring in business.

5. Suppose we classify 315 randomly selected college students according to their general major field and their self-described political viewpoint. The table presents the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Sciences** | **Business** | **Humanities** | **Social Sciences** |
| **Liberal** | 17 | 12 | 32 | 30 |
| **Moderate** | 33 | 40 | 23 | 20 |
| **Conservative** | 35 | 38 | 17 | 18 |

What percentage of liberals surveyed were humanities majors?

A. 44.4%

AR. Incorrect. You have used the conditional distribution for the viewpoint category “liberal” among students majoring in humanities. You actually need to use the conditional distribution for the major field category “humanities” among students describing themselves as liberals.

\*B. 35.2%

BR. Correct. We want to find the conditional distribution of the major field category “humanities” among students describing themselves as liberals. There are 17 + 12 + 32 + 30 = 91 liberals, of whom 32 are humanities majors. The correct result is therefore 32/91 = 0.352 or 35.2%.

C. 28.9%

CR. Incorrect. You have calculated the marginal distribution of the viewpoint category “liberal.” In this case, we must use a conditional distribution rather than a marginal distribution, because we are assuming that the person is a liberal.

6. Suppose we classify 315 randomly selected college students according to their general major field and their self-described political viewpoint. The table presents the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Sciences** | **Business** | **Humanities** | **Social Sciences** |
| **Liberal** | 17 | 12 | 32 | 30 |
| **Moderate** | 33 | 40 | 23 | 20 |
| **Conservative** | 35 | 38 | 17 | 18 |

Which of the following characteristics of these data supports the conclusion that there is an association between political viewpoint and general major field?

A. The sample does not contain equal number of individuals in each cell.

AR. Incorrect. Few categorical variables are so equally distributed, and such a distribution would actually represent a *lack* of association!

\*B. A higher proportion of students with a liberal political viewpoint major in the humanities, and a higher proportion of moderates and conservatives major in business.

BR. Correct. We say an association exists between two variables if specific values of one variable tend to occur with specific values of the other.

C. The marginal distributions of the two variables are not proportional.

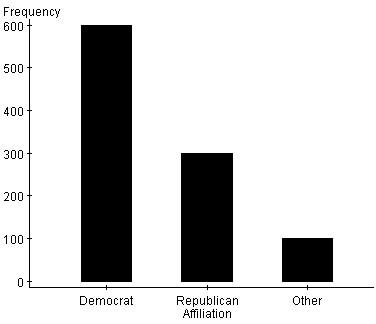
CR. Incorrect. The existence of an association involves examination of conditional distributions, not marginal distributions.

7. A review of voter registration records in a small town yielded the following table of the number of males and females registered as Democrat, Republican, or some other party affiliation.

|  |  |  |
| --- | --- | --- |
|  | **Male** | **Female** |
| **Democrat** | 300 | 600 |
| **Republican** | 500 | 300 |
| **Other affiliation** | 200 | 100 |

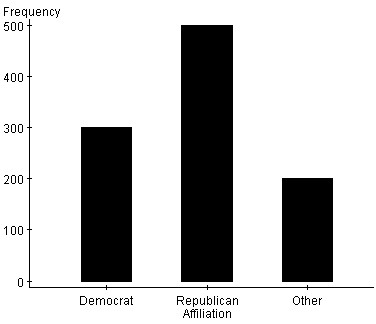
Which of the following bar graphs represents the distribution of party affiliation among all voters in this town?

A.



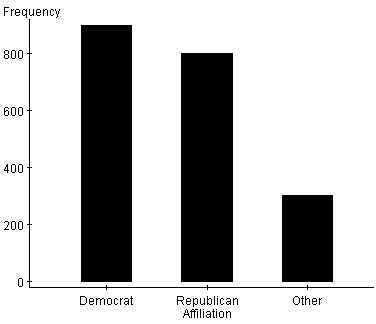
AR. Incorrect. This graph displays the distribution of party affiliation for females only.

B.



BR. Incorrect. This graph displays the distribution of party affiliation for males only.

\*C.



CR. Correct. There are 300 + 600 = 900 Democrats, 500 + 300 = 800 Republicans, and 200 + 100 = 300 Other in the set of all voters.

8. Below is a graphical presentation of information from the World Bank about the percentage of land in four East African countries that is forest.

Ethiopia

Kenya

Somalia

10

20

Percentage of land area



Eritrea



What’s wrong with this method of presenting information?

A. The vertical scale underrepresents differences between countries.

AR. Incorrect. Since the vertical scale starts at zero, there is no distortion of relative “tree” heights.

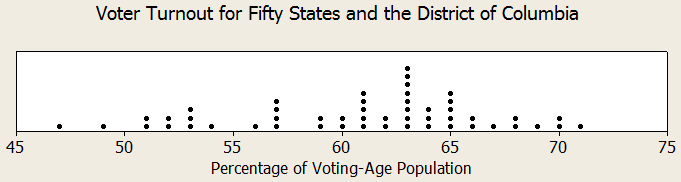
B. “Percentage of land area” does not take into account differences on total area among these countries.

BR. Incorrect. While it is true that total land area is not taken into account, it is probably more meaningful to compare relative abundance of forested land rather that total forested land.

\*C. Using proportionally-sized trees exaggerates differences between countries.

CR. Correct. The area of the tree for Eritrea is about nine times the area of the Kenya tree, but Eritrea’s forested land is only about three times that of Kenya.

9. Below is a dotplot of voter turnout (as a percentage of voting-age population) for the 50 states—plus the District of Columbia—in a recent presidential election.



Which of the following best describes this distribution?

\*A. Skewed slightly left, centered at about 62%, with a range of 24 percentage points.

AR. Correct. Skewed left means the “tail” is longer on the left side, and range is maximum – minimum—in this case 71 – 47 = 24. The median value is 62%.

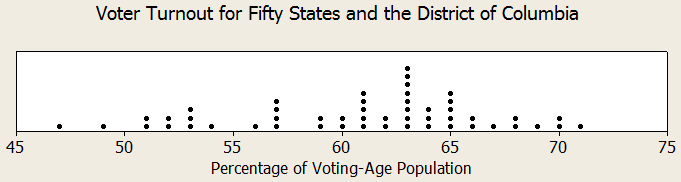
B. Roughly symmetric, centered at about 60%, with a range of 30 percentage points.

BR. Incorrect. The left tail is distinctly longer than the right, the median value is 62, and the range is 71 – 47 = 24, not 30.

C. Skewed slightly right, centered at about 62%, with a range of 24 percentage points.

CR. Incorrect. Skewed right means the “tail” is longer on the right side, not the left. The range and center are correct.

10. Below is a dotplot of voter turnout (as a percentage of voting-age population) for the 50 states—plus the District of Columbia—in a recent presidential election.



Approximately what percentage of the states had a turnout above 65%?

A. 27.5%

AR. Incorrect. This would be the result if you included the states with turnouts equal to 65%.

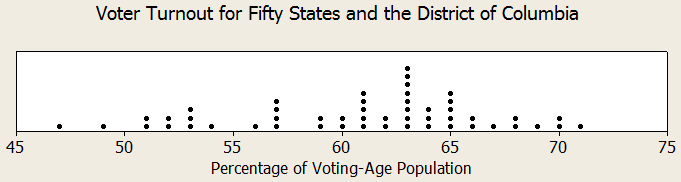
\*B. 17.6%

BR. Correct. There are 9 states with turnouts above 65%. 9/51 = 0.176 or 17.6%

C. 18%

CR. Incorrect. Be careful! Did you remember to use a denominator of 51? The data includes the District of Columbia!

11. Below is a dotplot of voter turnout (as a percentage of voting-age population) for the 50 states—plus the District of Columbia—in a recent presidential election.



A. 57 to 65

AR. The interquartile range is the difference between Q3 (65) and Q1 (57), not the interval itself.

B. 61

BR. Incorrect. This is the average of Q1 (57) and Q3 (65)

\*C. 8

CR. Correct. IQR = Q3 – Q1 or 65 – 57 = 8.

12. The stem-and-leaf diagram below gives the distribution of the ages in years of 20 participants at a family reunion.

|  |  |
| --- | --- |
| 0 | 7888 |
| 1 | 23689 |
| 2 | 45799 |
| 3 | 9 |
| 4 |  |
| 5 | 8 |
| 6 | 5 |
| 7 | 89 |
| 8 |  |
| 9 | 1 |

Which of the following statements about the distribution is correct?

\*A. The mean is larger than the median.

AR. Correct. Extremely large values, such as the ones seen in the right tail of this distribution, tend to pull the mean above the median. The mean, unlike the median, is a nonresistant measure of center.

B. The distribution is strongly skewed to the left.

BR. Incorrect. The distribution has a peak on the left side (at the 1 and 2 stems) and a long right tail, which means that it is skewed to the right. If it were skewed to the left, then the peak would be on the right side, at the higher stem values.

C. It makes the most sense to use the mean and standard deviation as a numerical summary of the center and spread of this distribution.

CR. Incorrect. The extremely large values in the right tail will strongly affect the mean and standard deviation, which are nonresistant measures of center and spread, respectively. The mean and standard deviation are not the best measures of center and spread to use here. The five-number summary would be a better choice.

13. Independently selected groups of middle-school children were given a poem to memorize. After a certain period of time, they were asked to recall as much of the poem as they could. A back-to-back stemplot of the distribution of the number of words that each group of children could correctly remember is displayed below.

|  |  |  |
| --- | --- | --- |
| **Group 1** |  | **Group 2** |
|  | 1 | 9 |
|  | 2 |  |
|  | 2 | 8 |
|  | 3 | 1 |
| 7 | 3 | 4 |
| 31 | 4 | 02333 |
| 8 | 4 | 668 |
| 44310 | 5 | 3 |
| 755 | 5 |  |
| 0 | 6 |  |

Which of the following statements about these data is true?

A. There are more students in Group 1 than in Group 2.

AR. Incorrect. There are thirteen students in each group.

B. In general, children in Group 2 were able to recall more words than children in Group 1.

BR. Incorrect. The peak of the Group 1 distribution is located at the first 5 stem, whereas the peak of the Group 2 distribution is located at the first 4 stem. Also, the median and mean of the Group 2 data are both larger than the corresponding measures for the Group 1 data. Children in Group 1—not Group2—appear to be able to recall more words, on average.

\*C. The third quartile of the Group 1 distribution is larger than the maximum value of the Group 2 distribution (that is, 25% of the Group 1 values are larger than any Group 2 value).

CR. Correct. The third quartile of the Group 1 distribution is located halfway between the third and fourth values in the ordered set of observations that are larger than the median, that is, at (55 + 55)/2 = 55. This is indeed larger than 53, the maximum value in the Group 2 distribution.

14. The following box plots show the distribution of test scores in Ms. Williams’s two sections of calculus.

score

70

80

90

100

60

50

A period

C period

Based on these box plots, which one of the following statements must be true?

\*A. The first quartile for the two sections is the same.

AR. Correct. For both sections is at or close to 72.

B. The lowest score in A period was close to the 15th percentile of the C period class.

BR. Incorrect. The graph tells us the 25th percentile (first quartile) and the minimum of the C period class, but there is no way to know where between those two values the 15th is located. See pages 60-65 in the text.

C. The mean score for students in C period was higher than the mean score for students in A period.

CR. Incorrect. The box plot is based on the five-number summary, which does not include the mean. It’s not possible to accurately locate the mean from the box plot.

15. In which of the following situations would a pie chart be an appropriate graph to use to summarize your data?

\*A. You want to display the distribution of favorite color for the students in your statistics class.

AR. Correct. You want to show how a categorical variable is distributed in a single group, so a pie chart (or a bar graph) is appropriate.

B. You want to compare the percentages of students in each grade at your school who favor a certain candidate for school president.

BC. Incorrect. Since the percentages you are comparing are from different groups rather than percentages rom the same “whole”, you can’t display the information in a pie chart. You need to use a bar graph.

C. You want to compare the life expectancies of different professions by displaying and comparing graphs of ages at death for random samples of famous scientists, authors, actors, and politicians.

CR. Incorrect. Age at death is a quantitative variable. To compare more than two distributions of quantitative data, a side-by-side boxplot would be the most appropriate graph.

16. The owner of a convenience store keeps track of how many customer buy lunch food during the “noon rush” each day for four days and calculates that the mean number of customers for those four days is 52. How many customers must come in on the fifth day to make the five-day mean 54?

A. 54

AR. Incorrect. Think about what the total number of customers must be for all five days in order for the mean to be 54.

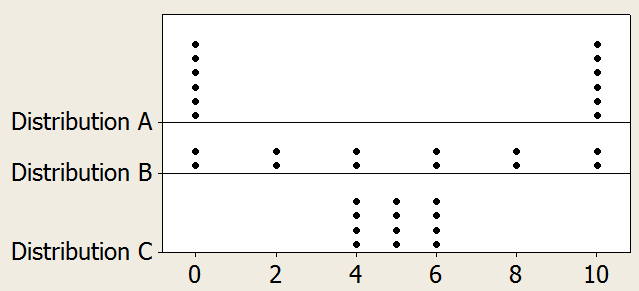
B. 60

BR. Incorrect. Think about what the total number of customers must be for all five days in order for the mean to be 54.

\*C. 62

CR. Correct. If the mean for five days is to be 54, the total number of customers for the entire period must be 5 x 54 = 270. There were 4 x 52 = 208 customers on the first four days, so the will have to be 62 on the fifth day to raise the mean to 54.

17. Which of the distributions displayed in the dot plots below has the highest standard deviation?



\*A. Distribution A

AR. Correct. Since standard deviation measures “typical distance” away from the mean of 5, the points in this distribution are all as far or farther from the mean as the points in the other two distributions.

B. Distribution B.

BR. Incorrect. Remember that standard deviation measures the “typical distance” of each data point from the mean. Some of these points are equal or very close to the mean, which is not the case with one of the other distributions.

C. Distribution C

CR. Incorrect. Remember that standard deviation measures the “typical distance” of each data point from the mean. All of these points are equal or very close to the mean, which is not the case with one of the other distributions.

18. According to the 1.5 × IQR rule, how many outliers are there in the data set 72, 110, 114, 115, 118, 123, 144, 156?

A. None.

AR. Incorrect. Since Q1 = 112 and Q3 = 133.5, the interquartile range is 133.5 – 112 = 21.5. A value is considered to be an outlier if it lies below Q1 – (1.5)(IQR) or above Q3 + (1.5)(IQR). At least one value satisfies this requirement.

\*B. One.

BR. Correct. Since Q1 = 112 and Q3 = 133.5, the interquartile range is 133.5 – 112 = 21.5. A value is considered to be an outlier if it lies below 112 – (1.5)(21.5) = 79.75 or above 133.5 + (1.5)(21.5) = 165.75. The only value meeting this requirement is 72.

C. Two.

CR. Incorrect. Since Q1 = 112 and Q3 = 133.5, the interquartile range is 133.5 – 112 = 21.5. A value is considered to be an outlier if it lies below Q1 – (1.5)(IQR) or above Q3 + (1.5)(IQR). There are less than two values that meet this requirement.

19. The histogram below shows the typical travel time to school (self-reported) for 50 high school students.



Which of the following statements is true about the mean and median of this distribution?

\*A. Mean > Median

AR. Correct. The distribution is skewed right, which means that the very large values on the right tail contribute disproportionally to the mean, making it higher than the median.

B. Mean < Median

BR. Incorrect. The distribution is skewed right, which means that the very large values on the right tail contribute disproportionally to the mean, making it higher than the median.

C. Mean = Median

CR. Incorrect. The distribution is skewed right, which means that the very large values on the right tail contribute disproportionally to the mean, making it higher than the median.

20. Jerome’s summer reading list has 8 books, and he is examining the number of pages in each book. After calculating the mean, median, standard deviation, and interquartile range, he realized that the longest book is actually 100 pages longer than he thought it was. Which of his measurements does he need to recalculate?

A. The mean, standard deviation, and interquartile range.

AR. Incorrect. The interquartile range is resistant to extreme values: neither Q1 nor Q3 will change.

\*B. The mean and standard deviation.

BR. Correct. The mean and standard deviation are both sensitive to outliers, but the median and interquartile range are resistant.

C. Only the mean.

CR. Incorrect. While the mean is indeed sensitive to outliers, so is the standard deviation.