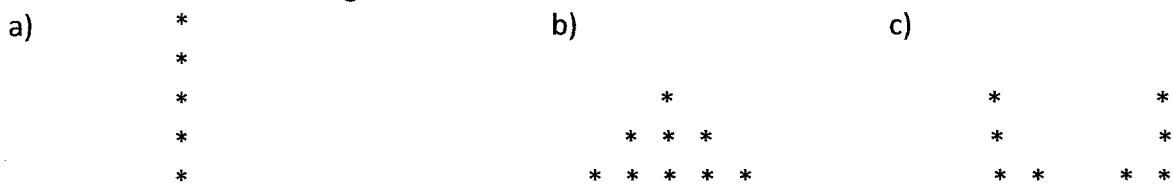


- 1) In a random sample of 300 registered Florida voters, 23% are found to be registered as a party other than Republican or Democrat. Identify the population and the sample.
- 2) For the problem above, is the figure of 23% a statistic or a parameter?
- 3) Inferential statistics involves using information from a sample to draw conclusions about a population. We discussed several of its advantages (savings in time, money, etc.), but what is its disadvantage?
- 4) Classify the data below as either qualitative or quantitative:
  - a) Political party affiliation of voters: Democratic, Republican, Independent, Libertarian, Green, etc.
  - b) Ages of students enrolled in a class: 19, 22, 23, 20, 34, 45, 18, 27, etc.
- 5) State the level of measurement (nominal, ordinal, interval, or ratio) for the data below:
  - a) Drink size: small, medium, large
  - b) High temperature (Fahrenheit): 87, 90, 95, 78
  - c) Occupation: teacher, secretary, surgeon
  - d) Number of students enrolled in STA2023 classes: 23, 30, 20, 15, 27
- 6) A group of 60 students take an S.A.T. math test (without calculators) and score an average of 480. The same students retake the S.A.T. math test a month later (with calculators) and score an average of 525. Give two reasons why attributing the rise solely to the use of calculators is unwise, and suggest a better way to set up the experiment.
- 7) To test the effectiveness of a drug to help people quit smoking, briefly explain why a randomized block design (where subjects are blocked according to how long they have been smoking) would be better than a completely randomized design.
- 8) Identify the sampling method (simple random, stratified, cluster, systematic, or convenience) used in the examples below:
  - a) A large city's population is 48% male and 52% female. So to obtain a sample of 200 people we randomly sample 96 of the males and then randomly sample 104 of the females.
  - b) To obtain a sample of 50 shoppers at a mall we use the first 50 shoppers who are willing to speak with us.
  - c) To obtain a sample of 12 students from a class of 300 (students listed from 1 to 300), we use a calculator to generate 12 random integers from 1 to 300 and then use the corresponding 12 students.
  - d) A school has a total of 2000 students (80 classrooms with 25 students each). To obtain a sample of 100 students we randomly select 4 of the classrooms and use all students in the selected classrooms.
  - e) We want a sample of 30 students from a large university. From a list of students (in alphabetical order), we randomly select the 23<sup>rd</sup> student and then every 50<sup>th</sup> student thereafter until we have 30 students.
- 9) Using 5 classes, construct a frequency distribution (and relative frequency histogram) for the data set below:  
 182 163 214 159 206 167 153 221 174 200 193 167 169 178
- 10) Construct a stem-and-leaf plot for each data set below (use the indicated number of stems):
  - a) 3.12 3.28 2.92 2.90 2.83 (5 stems)
  - b) 2.5 6.4 4.2 4.5 5.1 (6 stems)
  - c) 72 65 68 61 57 53 (6 stems)

3.17 3.21 3.07 3.02 2.88                      5.4 7.0 4.2 5.8 3.3                      76 62 72 77 70 68
- 11) Construct a dot plot for the data set below and state the term that best describes its distribution shape:  
 4        5        9        3        5        5        7        8        4        6        4        6        3
- 12) Compute the median and range for the data set 8, 10, -2, 1, 12, 3
- 13) Compute the mean, median, mode, variance, and standard deviation for the sample below:  
 15, 18, 19, 20, 22, 22, 24, 26, 28, 30, 31, 35, 39, 41, 48
- 14) Is it possible to have a data set for which mode < mean < median?  
 If so, give an example of such a data set using five observations.
- 15) If one data set has a larger range than another, must it also have a larger variance? Explain.
- 16) The graphs of three data sets are shown below. If the axes are scaled in the same way for each data set, which data set has the largest variance?



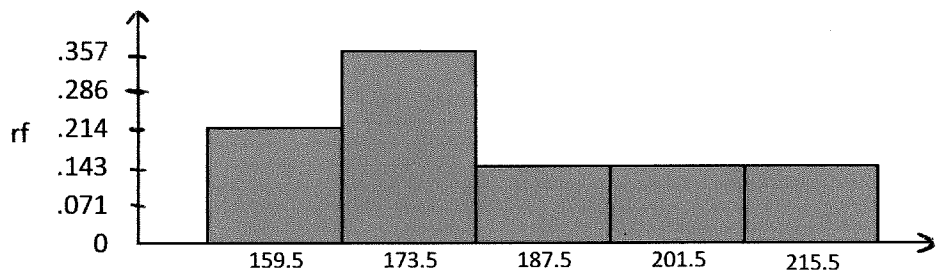
- 17) if a data set is normally distributed with a mean of 124 and a standard deviation of 16, what proportion of the data is ...
- a) between 108 and 140?      b) less than 108?      c) greater than 156?
- 18) If a data set has a mean of 64 and a standard deviation of 12, but its distribution shape is unknown, what proportion of the data is ...
- a) between 40 and 88?      b) between 28 and 100?      c) greater than 100?
- 19) If a test taker's score is at the 85<sup>th</sup> percentile it must mean that ...
- a) their score is lower than 85% of the other scores  
 b) their score is 85 points higher than the average score  
 c) they answered 85% of the test questions correctly  
 d) their score is higher than 15% of the other scores  
 e) their score is higher than 85% of the other scores
- 20) A particular species of fish has a mean length of 18.5 inches and a standard deviation of 1.5 inches. Find z-scores for the following lengths:
- a) 21.2 inches      b) 18.5 inches      c) 17.6 inches
- 21) For the problem above, find the lengths of fish with the following z-scores:
- a) 0.4      b) 0      c) -2.4

**Answers:**

- 1) population: all registered Florida voters      sample: the 300 voters  
 2) statistic (since it comes from only a sample)  
 3) from a sample we can obtain only an estimate about the population, and an estimate is subject to error.  
 4) a) qualitative    b) quantitative  
 5) a) ordinal    b) interval    c) nominal    d) ratio  
 6) The scores may have risen simply because the students were more familiar with the test the second time around or because they engaged in additional study during the month between tests.  
 It would have been better to have two different groups of students take the test (students of similar average knowledge levels)  
 Group 1: without calculators and then with calculators  
 Group 2: without calculators on both tests  
 This way if group 1 showed greater improvement than group 2 on the retake, it would be safer to attribute the difference to the use of the calculators since BOTH groups will have had more familiarity with the test the second time around and BOTH groups will have had the chance for more study between tests.  
 7) If a person has been smoking for decades (instead of just months), it would be much harder for that person to quit smoking. So blocking according to a person's years of smoking would improve the study.  
 8) a) stratified    b) convenience    c) simple random    d) cluster    e) systematic

9)

class	f	rf
153-166	3	$3/14 \approx .214$
167-180	5	$5/14 \approx .357$
181-194	2	$2/14 \approx .143$
195-208	2	$2/14 \approx .143$
209-222	2	$2/14 \approx .143$
	14	1



- 10) a)  $28 \overline{)38}$       b)  $2 \overline{)5}$       c)  $5 \overline{)3}$
- $29 \overline{)02}$        $3 \overline{)3}$        $5 \overline{)7}$
- $30 \overline{)27}$        $4 \overline{)225}$        $6 \overline{)12}$
- $31 \overline{)27}$        $5 \overline{)148}$        $6 \overline{)588}$
- $32 \overline{)18}$        $6 \overline{)4}$        $7 \overline{)022}$
- $7 \overline{)0}$        $7 \overline{)67}$

- 11)    • •    shape:    12) median=5.5, range=14  
 • • • •    skewed right  
 • • • • •  
 3 4 5 6 7 8 9

- 13) mean=27.87    median=26    mode=22  
 variance=89.84    standard deviation=9.48  
 14) Yes: 1,1,5,6,7    mode=1, mean=4, median=5,  
 1 < 4 < 5

key:  $28 \overline{)3} = 2.83$

key:  $2 \overline{)5} = 2.5$     key:  $5 \overline{)3} = 53$

- 15) No. For example: 0, 50, 50, 50, 100 (larger range)  
 1, 2, 50, 98, 99 (larger variance)
- 16) c
- 17) a)  $\approx 68\%$     18) a) at least 75%    19) e    20) a) 1.8    21) a) 19.1 inches  
 b)  $\approx 16\%$     b) at least 89%    b) 0    b) 18.5 inches  
 c)  $\approx 2.5\%$     c) at most 11%    c) -0.6    c) 14.9 inches