Material covered: Course Pack: Lecture Unit 1, Lecture Unit 2, Lecture Unit 3

WARNING: these practice problems may not cover all the material for which you are responsible!

1. The heights of American men aged 18 to 24 are approximately symmetric and mound shaped with mean 68 inches and standard deviation 2.5 inches. Half of all young men are shorter than
   a) 65.5 inches  b) 68 inches  c) 70.5 inches  d) can’t tell because the median height is not given

2. Use the information in Problem 1. What percent of young men are taller than 6’ 1”?
   a) 94.5%    b) 44.5%    c) 5.5%    d) 2.5%

3. A Consumer Reports article on energy bars gave the brand name, flavor, price, number of calories, and grams of protein and fat. Describe the Who, What, Why, Where, When, How. What variables are categorical? What variables are quantitative?

4. Suppose for a set of data \( x_1, x_2, \ldots, x_7 \) we calculate \( \sum_{i=1}^{n}(x_i - \bar{x})^2 = 4.51 \), then the standard deviation is
   a) .75    b) .87    c) .64    d) .80    e) none of the above

5. A standardized test designed to measure math anxiety has a mean of 100 and a standard deviation of 10 in the population of first year college students. Which of the following observations would you suspect is an outlier?
   a) 150    b) 100    c) 90    d) all of the above    e) none of the above

6. Which of the following best describes a risk in using the sample range to measure spread? Choose one of the following:
   a) The statistic is not in the same unit of measurement as the observations themselves.
   b) The largest or smallest observation may be a mistake or an outlier.
   c) It is complicated to compute, which may result in an error.
   d) It always produces spreads that are too large.

7. The following 2-way table shows students by major and home state for a small private school in the northeast U.S.

<table>
<thead>
<tr>
<th>Home State</th>
<th>Biology</th>
<th>Accounting</th>
<th>History</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>80</td>
<td>65</td>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>NJ</td>
<td>50</td>
<td>40</td>
<td>65</td>
<td>95</td>
</tr>
<tr>
<td>NY</td>
<td>75</td>
<td>50</td>
<td>45</td>
<td>80</td>
</tr>
<tr>
<td>MD</td>
<td>65</td>
<td>55</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

   a. Find the marginal distribution for home state.
   b. Find the conditional distribution (in percentages) of major distribution for the home state of NJ.
   c. Find the conditional distribution (in percentages) of home state distribution for the biology major.

8. Twenty-eight applicants interested in working for the Food Stamp program took an examination designed to measure their aptitude for social work. The following test scores were obtained:
   79, 93, 84, 86, 77, 63, 46, 97, 87, 88, 87, 92, 68, 72, 86, 98, 81, 70, 66, 98, 59, 76, 68, 91, 94, 85, 88, 79.
   a. Find \( Q_1 \).
   b. Construct a boxplot for these observations. Do you observe any outliers?

9. Two students are enrolled in an introductory statistics course at the University of Florida. The first student is in a morning section and the second student is in an afternoon section. If the student in the morning
section takes a midterm and earns a score of 76, while the student in the afternoon section takes a midterm with a score of 72, which student has performed better compared to the rest of the students in his respective class? For the morning class, the class mean was 64 with a standard deviation of 8. For the afternoon class, the class mean was 60 with a standard deviation of 7.5.

10. Raw scores on evaluation tests given to elementary school children are often transformed for easier comparison. A test of reading ability has mean 72 and standard deviation 25 when given to third graders. So that we can compare third grade scores with “normed” scores of students in other grades, we want scores in the third grade to have mean 100 and standard deviation 20.
   a. What linear transformation will change third grade scores $x$ into new scores $x_{new} = a + bx$ that have the desired mean and standard deviation? (Use $b > 0$ to preserve the order of the scores).
   b. Nancy is a third grade student who scored 75 on the test. Find Nancy's transformed score.

11. Which of the following describes a basic DIFFERENCE between nominal data and ordinal data?
   a. Nominal and ordinal data are both quantitative data.
   b. Nominal and ordinal data are both qualitative data.
   c. Nominal data are qualitative data whereas ordinal data are quantitative data.
   d. Nominal data have a meaningful zero point and the ratio of the two data values is meaningful whereas with ordinal data the zero point is arbitrary.
   e. Nominal data classify items into one of a set of categories whereas ordinal data classify items into categories that can be ranked and ordered.

12. In the stem and leaf plot below the stems are the one's digit and the leaves are the first digit to the right of the decimal point, so 2 | 1 denotes the value 2.1

   2 | 1 2 5 8 9
   3 | 3 4 5 8
   4 | 2 2 4 5
   5 | 1 5
   6 | 1 5
   7 | 8
   8 | 1

   a. Is the data right-skewed or left-skewed?
   b. What is the median?
   c. What is $Q_3$?

13. A manufacturer of computer hard disk drives knows that its computer hard disk drives have a mean time to failure of 24 months with a standard deviation of 4 months. The failure data is known to be symmetric and mound-shaped. Determine the following:
   a. What is the approximate proportion of the manufacturer's hard drives that will fail later than 16 months?
   b. What is the approximate proportion of the manufacturer's hard drives that will fail between 16 months and 28 months?
   c. If the manufacturer offers a 20 month warranty on its hard drive, what is the approximate proportion of these hard drives that could fail under warranty?

14. The following data are a sample of the percentage rates of return experienced by six-month Eurodollar deposit purchasers:

   $-1 \ 2 \ -2 \ 1 \ -1 \ 4 \ 2 \ 3 \ 1$

   The standard deviation is:
   a. 4   b. 2   c. 1.05   d. 1.11   e. can't be done since some values are negative
Consider the following histograms of variables labeled X1, X2 and X3:

15. The median for variable X2 would be around
   a. 10     b. 305     c. 250     d. impossible to tell

16. The third quartile for variable X1 would be around
   a. 12     b. 8     c. 5     d. 15

17. The distribution in which the mean and median are most different would be
   a. X1     b. X2     c. X3     d. It is impossible to tell.

18. The standard deviation for variable X1 would be
   a. About the same as the standard deviation for variable X2.
   b. Smaller than the standard deviation for variable X2.
   c. Larger than the standard deviation for variable X2.
   d. It is impossible to tell.

19. The histograms above are the results of questions asked of a group of undergraduate students. Match the histogram (X1, X2, or X3) above to the appropriate question below.
   a. How many hours did you work at a job last week? ______
   b. What is your shoe size? ______
   c. How much did you spend on textbooks (in dollars) this semester? ______

20. A computer manufacturer laid off 50 of its 100 software engineers. The ages of the engineers not laid off and the ages of the engineers laid off are shown graphically in the boxplots below. Use the boxplots to determine if each of the following statements are True or False
   a. The median age of the laid-off employees and the median age of those not laid off are about the same.
   b. The IQR of the ages of laid-off employees is greater than the IQR of the ages of employees not laid off.
   c. 25% of the laid off employees are older than all but one of the employees not laid off.
Questions 21, 22, and 23 refer to the following:

A marketing research firm wishes to determine if the adult men in Laramie, Wyoming would be interested in a new upscale men's clothing store. From a list of all residential addresses in Laramie, the firm selects a simple random sample of 100 and mails a brief questionnaire to each.

21. In the above survey, what is the sample?
   a. all adult men in Laramie, Wyoming
   b. all residential addresses in Laramie, Wyoming
   c. the members of the marketing firm that actually conducted the survey
   d. the adult men at the 100 addresses the survey was mailed to

22. In the above survey, what is the population of interest?
   a. all adult men in Laramie, Wyoming
   b. all residential addresses in Laramie, Wyoming
   c. the members of the marketing firm that actually conducted the survey
   d. the 100 addresses the survey was mailed to

23. In the above survey, what is the chance that all 100 homes in a particular neighborhood in Laramie end up being the sample of residential addresses selected?
   a. the same as for any other set of 100 residential addresses
   b. exactly 0 since random samples will spread out the addresses selected
   c. reasonably large due to the “cluster” effect
   d. 100 divided by the size of the population of Laramie

24. Which of the following is not a potential source of bias in a survey?
   a. undercoverage
   b. nonresponse
   c. the use of convenience sampling
   d. the use of simple random sampling
25. The January 2005 Gallup Youth Survey telephoned a random sample of 1,028 U.S. teens aged 13-17 and asked these teens to name their favorite movie from 2004. Napoleon Dynamite (vote for Pedro! he has skills!) had the highest percentage with 8% of teens ranking it as their favorite movie.

Which of the following statements are true?
I. The population of interest is U.S. teens aged 13-17.
II. 8% is a statistic and not the actual percentage of all U.S. teens who would rank this movie as their favorite.
III. This sampling design should provide a reasonably accurate estimate of the actual percentage of all U.S. teens who would rank this movie as their favorite.

a. I only b. II only c. III only d. I and II e. I, II, and III

26. Suppose your local school district decides to randomly test high school students for attention deficit disorder (ADD). There are three high schools in the district, each with grades 9-12. The school board pools all of the students together and randomly samples 250 students. Is this sample a simple random sample?

a. Yes, because the students were chosen at random.
b. Yes, because each student is equally likely to be chosen.
c. Yes, because they could have chosen any group of 250 students from throughout the school district.
d. No, because we cannot guarantee that there are students from each school in the sample.
e. No, because we cannot guarantee that there are students from each grade in the sample.

27. A chemistry professor who teaches a large lecture class one day surveys the students in attendance about how he can make the class more interesting, hoping he can get more students to attend lectures. This method suffers from

a. voluntary response bias  b. nonresponse bias  c. response bias
d. undercoverage  e. none of the above

28. When every possible sample with the same number of observations is equally likely to be chosen, the selected sample is called a:

a. simple random sample.
b. stratified random sample
c. cluster random sample
d. systematic random sample

29. When the population is divided into mutually exclusive sets, and then a simple random sample is drawn from each set, this is called:

a. simple random sampling.
b. stratified random sampling.
c. cluster random sampling.
d. systematic random sampling.

30. A marketing research firm divides the population of a state into geographic areas, and randomly selects some of the areas and takes a simple random sample of each selected area. This is an example of a

a. cluster random sample
b. systematic random sample
c. simple random sample
d. stratified random sample

31. A simple random sample of 20 undergraduates at Johns Hopkins University found that 60% of those sampled felt that drinking was a problem among college students. A simple random sample of 20 undergraduates at Ohio State University found that 70% felt that drinking was a problem among college students. The number of undergraduates at Johns Hopkins University is approximately 2,000; the number at Ohio State is approximately 50,000. We conclude

a. the sample from Johns Hopkins is much less representative of its population than the sample from Ohio State is of its population.
b. the sample from Johns Hopkins is much more representative of its population than the sample from Ohio State is of its population.
c. the samples from Johns Hopkins and Ohio State are equally representative of their respective populations.
d. it is impossible to make any statements about which sample is more representative of its corresponding population since the students surveyed attended different schools.

32. A simple random sample of 2% of the undergraduates at Johns Hopkins University found that 60% of those sampled felt that drinking was a problem among college students. A simple random sample of 2% of the undergraduates at Ohio State University found that 70% felt that drinking was a problem among college students. The number of undergraduates at Johns Hopkins University is approximately 2,000; the number at Ohio State is approximately 50,000. We conclude
a. the sample from Johns Hopkins is much less representative of its population than the sample from Ohio State is of its population.
b. the sample from Johns Hopkins is much more representative of its population than the sample from Ohio State is of its population.
c. the samples from Johns Hopkins and Ohio State are equally representative of their respective populations.
d. it is impossible to make any statements about which sample is more representative of its corresponding population since the students surveyed attended different schools.

ANSWERS

1. b d (73 inches is how many standard deviations above the mean?)


4. b 5. a

6. The largest or smallest observation may be a mistake or an outlier.

7. a. Home State Number of Students (%)
   PA 300 (300/1000 = 30%)
   NJ 250 (250/1000 = 25%)
   NY 250 (250/1000 = 25%)
   MD 200 (200/1000 = 20%)

b. Major Conditional for NJ
   Bio 20%
   Acct. 16%
   Hist 26%
   Educ 38%

c. Home State Conditional for Biology
   PA 29.6%
   NJ 18.5%
   NY 27.8%
   MD 24.1%

8. a. First, order the data from smallest to largest. Since we want $Q_1$ and there are $n=28$ observations, $Q_1$ is the median of the smallest 14 data values, that is, the mean of the data values in ordered positions 7 and 8:
   
   $$Q_1 = \frac{70 + 71}{2} = 71.$$
No outliers are observed.

9. \( z_1 = (76-64)/8 = 1.5 \); \( z_2 = (72-60)/7.5 = 1.6 \). The student in the afternoon section performed better.

10. a. 2 equations, 2 unknowns \( a \) and \( b \); (i) new mean must be 100, so we have the equation \( a + b(72) = 100 \); (ii) new standard deviation must be 20 so we have the equation \( b(25) = 20 \). Equation (ii) \( \Rightarrow b = \frac{20}{25} = 0.8 \); substitute \( b = 0.8 \) into equation (i) to get \( a = 100 - 0.8(72) = 100 - 57.6 = 42.4 \); so the transformation is \( x_{new} = 42.4 + 0.8x \).

b. \( x_{new} = 42.4 + 0.8(75) = 102.4 \)

11. e. 12. a. right-skewed; b. \( m = \frac{3.8+4.2}{2} = 4.0 \); \( Q_3 = \frac{5.1+5.5}{2} = 5.3 \). 13. (use 68-95-99.7 rule) a. \( .975 \)
b. \( .815 \) c. \( .16 \) 14. b 15. b 16. a 17. c 18. b
19. a. X3 b. X1 c. X2 20. a) T b) F c) T d) T
21. d 22. a 23. a 24. d 25. e 26. c 27. d (the students not in attendance are not included in the survey)
28. a 29. b 30. a 31. c 32. a