1. Which is true of the data shown in the histogram?
   I. The distribution is skewed to the right.
   II. The mean is smaller than the median.
   III. We should use the median and IQR to summarize these data.

   A. I only
   B. II only
   C. III only
   D. II and III only
   E. I, II, and III

2. The five-number summary of credit hours for 24 students in an introductory statistics class is:

<table>
<thead>
<tr>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.0</td>
<td>15.0</td>
<td>16.5</td>
<td>18.0</td>
<td>22.0</td>
</tr>
</tbody>
</table>

   From this we know that
   A. there are no outliers in the data.
   B. there is at least one low outlier in the data.
   C. there is at least one high outlier in the data.
   D. None of the above.

3. Suppose that a Normal model describes fuel economy (miles per gallon) for automobiles and that a Saturn has a standardized (z-score) of +2.2. This means that Saturns . . .

   A. get 2.2 miles per gallon.
   B. get 2.2 miles per gallon more than the average car.
   C. have a standard deviation of 2.2.
   D. achieve fuel economy that is 2.2 standard deviations better than the average car.

4. Suppose a Normal model describes the number of pages printer ink cartridges last for. If we keep track of printed page counts for the 47 printers in the company's office, which must be true?
   I. The page counts of those ink cartridges will be normally distributed.
   II. The histogram for those page counts will be symmetric.
   III. 95\% of those page counts will be within 2 standard deviations of the mean.

   A. None
   B. I only
   C. II only
   D. II and III
   E. I, II, and III

5. A lakeside restaurant found the correlation between the daily temperature and the number of meals they served to be 0.40. On a day when the temperature is two standard deviations above the mean, the number of meals they should plan on serving is ______ the mean.

   A. equal to
   B. 0.4 SD above
   C. 0.8 SD above
   D. 2.0 SD above

6. For families who live in apartments the correlation between the family's income and the amount of rent they pay is \( r = 0.60 \). Which is true?
   I. In general, families with higher incomes pay more rent.
   II. On average, families spend 60\% of their income on rent.
   III. 60\% of the variation in rent is explained by a linear model.

   A. I only
   B. II only
   C. I and III only
   D. I, II, and III
7. Which scatterplot shows a strong association between two variables even though the correlation is probably near zero?

A.  
B.  
C.  
D.  

8. The residuals plot for a linear model is shown. Which is true?

A. The linear model is okay because approximately the same number of points are above the line as below it.
B. The linear model is okay because the association between the two variables is fairly good.
C. The linear model is no good because the correlation is near 0.
D. The linear model is no good because of the curve in the residuals.

9. A regression analysis of a company profits and the amount of money the company spent on advertising found $R^2 = 0.72$. Which of these is true?

I. This model can correctly predict the profit for 72% of the companies.
II. On average, about 72% of a company's profit results from advertising.
III. On average, companies spend about 72% of their profits on advertising.

A. None  
B. I only  
C. II only  
D. III only

10. Extrapolation is

A. okay to do as long as we are making predictions into the past, not the future.
B. okay to do if we tell people that we are assuming the linear relationship will hold outside the range of the data.
C. okay to do as long as there were no outliers in the original data.
D. not okay to do.

11. Two variables that are actually not related to each other may nonetheless have a very high correlation because they both result from some other, possibly hidden, factor. This is an example of

A. Extrapolation  
B. A lurking variable  
C. Influential points  
D. Leverage points

12. If the point in the upper right hand corner of the scatterplot is removed from the data set, then what will happen to the slope of the line of best fit ($b$) and the correlation ($r$)?

A. Both will increase.
B. Both will decrease.
C. $b$ will increase, $r$ will decrease.
D. $b$ will decrease, and $r$ will increase.
13. 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 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 250 students from throughout the district.
D. No, because we can’t guarantee that there are students from each school in the sample.
E. No, because we can’t guarantee that there are students from each grade in the sample.

14. In an intro stats class, 27% of students eat breakfast in the morning and 60% of students floss their teeth. Six percent of students eat breakfast and floss their teeth.

What is the probability that a student from this class neither eats breakfast nor flosses their teeth?

A. 87%  
B. 13%  
C. 19%  
D. 81%

15. A recent survey said that 60% of college students live on campus, 40% have a campus meal program, and 24% do both. Living on campus and having a meal plan are:

A. Independent  
B. Disjoint  
C. Both independent and disjoint  
D. Neither independent nor disjoint

16. A certain population is approximately Normal. We want to estimate its mean, so we will collect a sample. Which should be true if we use a large sample rather than a small one?

I. The distribution of our sample data will be approximately Normal.  
II. The sampling distribution of the sample means will be approximately Normal.  
III. The variability of the sample means will be smaller.

A. I only  
B. II only  
C. III only  
D. II and III  
E. I, II, and III

17. It is generally believed that electrical problems affect about 14% of new cars. An automobile mechanic conducts diagnostic tests on 128 new cars on the lot. How many successes (electrical failures) do you expect?

A. 1.28  
B. 17.92  
C. 110.08  
D. Cannot determine from information given.

18. Which is true about a 95% confidence interval for a population proportion based on a given sample? (Circle all that apply)

A. We are 95% confident that the sample proportion is in our interval.  
B. There is a 95% chance that our interval contains the population proportion.  
C. The interval is wider than a 99% confidence interval would be.  
D. If we repeat the experiment multiple times, then 95% of those 95% CI’s capture the true population parameter.
19. Name that test! For each scenario, determine which statistical test is appropriate, based on how the data is collected. Don’t worry if the assumptions are met, just write down the test.

a. We want to determine whether Amherst has (historically) a better basketball team than Williams. We randomly select 30 games from each team history and count the number of victories.

b. We want to determine whether Amherst has (historically) a better basketball team than Williams. We randomly select 30 games where they have played against each other and count the number of victories.

c. We want to determine whether Amherst has (historically) a better basketball team than Williams. We randomly select 30 games from each team history and record the number of points scored in each game.

d. We want to determine whether Amherst has (historically) a better basketball team than Williams. We randomly select 30 games where they have played each other and record the number of points scored in each game.
21. One last multiple choice: Suppose you want determine how a population mean compares to a conjectured mean of 5 units. You collect 20 observations and run a one sample t test, attaining a t score of −2.090. Assuming all assumptions are met, for which alternative hypotheses would you reject the null at a .05 level of significance? (Circle all that apply)

   A. $H_A: \mu > 5$
   B. $H_A: \mu < 5$
   C. $H_A: \mu \neq 5$
   D. None of the above

22. For questions a through c, assume a fair six-sided die (numbered 1 through 6) is rolled 5 times.

   a. What is the probability that exactly two 4’s appear in 5 rolls?

   b. What is the probability that the last two rolls are both 4’s?

   c. What is the probability that the only two 4’s that appear are on the last two rolls?
23. A nerd complains that the number of 4’s appearing on this die is more than what we would expect from a fair die. To quickly test this claim, we roll the die 100 times and see that the number 4 shows up 19 times. Test, to a .10 level of significance, whether the die is unfair (in the way the nerd complained). Clearly state:

- Which test you are using.
- Appropriate null and alternative hypotheses.
- Assumption checks (doing relevant calculations if needed).
- Calculation of test statistic.
- Whether or not you reject.
- Your conclusion in context.
Tear of this sheet for scratchwork for multiple choice, not to be turned in.