This worksheet reviews some of the basic concepts from the course. Your score on this worksheet will replace your lowest quiz/homework grade. Work this assignment in class, allowing yourself no more than 60 minutes. When you are finished, turn in this worksheet in my drop-box, right outside my office (Bagby 125).

Use the following scenario to answer question 1 through 5: How many movies does an adult American watch per year? An opinion poll calls 2000 randomly chosen residential telephone numbers, then asks to speak with an adult member of the household. The interviewer asks, “How many movies have you watched in a movie theatre in the past 12 months?”. In all 1131 people respond to this question. The results are in the following table.

<table>
<thead>
<tr>
<th>Number of movies seen in the last 12 months</th>
<th>0</th>
<th>1 to 3</th>
<th>4 to 8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>385</td>
<td>463</td>
<td>181</td>
<td>102</td>
</tr>
</tbody>
</table>

1. The type of data being asked for in this opinion poll is
   a) quantitative  b) explanatory  c) categorical  d) response  e) none of these

2. The population of interest in this survey is
   a) all adult Americans  b) all adult Americans who watch movies  c) the 2000 people called
d) the 1131 people who answered the question  e) none of these

3. The sample in this survey is
   a) all adult Americans  b) all adult Americans who watch movies  c) the 2000 people called
d) the 1131 people who answered the question  e) none of these

4. What is the non-response rate for this survey?
   Answer: ______________________

5. Based on this data, the polling organization estimates that 9.2% of adult Americans watched 9 or more movies in a theatre in the past 12 months, with a margin of error of ±1.7% for 95% confidence. What statistical procedure was used in making this estimate?
   a) 1 sample t test  b) 1 sample t interval  c) matched pairs t test  d) matched pairs t interval
e) 2 sample t test  f) 2 sample t interval  g) 1 proportion z test  h) 1 proportion z interval
i) 2 proportion z test  j) 2 proportion z interval
Use the following scenario to answer questions 6 and 7. Here is the probability model for the blood type of a randomly chosen person in the United States.

<table>
<thead>
<tr>
<th>Blood Type</th>
<th>O</th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.45</td>
<td>0.4</td>
<td>0.11</td>
<td>??</td>
</tr>
</tbody>
</table>

6. What is the probability that a randomly chosen person in the US has blood type AB?
Answer:____________________

7. What is the probability that a randomly chosen person has neither type O nor type AB blood?
Answer:____________________

Use the following scenario to answer questions 8 through 11. Scores on the SAT college entrance exam in a recent year were roughly normal, with a mean of 1075 and a standard deviation of 200.

8. Pick one student at random from this population. What is the probability that the chosen student’s SAT score is within ±10 points of the mean score 1075?
   a) 0.0399 b) 0.3829 c) 0.6171 d) 0.9601 e) none of these

9. Pick a simple random sample (SRS) of one hundred students from this population, and compute their average SAT score. If you repeated this over and over again, each time with a new random sample, what would the standard deviation of the average scores be?
   Answer:____________________

10. Jane makes a SAT score of $x = 1350$. What is Jane’s $z$-score?
   Answer:____________________

11. How large a sample size $n$ is necessary in order for $\sigma(\bar{x})$ to be no more than 12.5?
   a) $n = 4$ b) $n = 16$ c) $n = 128$ d) $n = 256$ e) none of these

12. Circle all of the items below that are measures of spread of a distribution.
   a) mean b) standard deviation c) median d) IQR e) none of these

13. Circle all of the items below that are measures of center of a distribution.
   a) mean b) standard deviation c) median d) IQR e) none of these

14. Circle all of the items below that are resistant to outliers.
   a) mean b) standard deviation c) median d) IQR e) none of these

15. The mean time spent waiting in line at the local Walmart is somewhat higher than the median time spent waiting in line. This means that the distribution of survival times is most likely
   a) left-skewed b) right-skewed c) symmetric d) unimodal e) none of these

16. If a distribution is strongly skewed, then the most appropriate numerical summary of the distribution is
   a) the two-number summary b) the five-number summary c) the correlation coefficient
d) the $\chi^2$-statistic e) none of these
17. The $P$-value for a hypothesis test is $P = 0.375$. At which of the following significance levels would we reject the null hypothesis?
   a) $\alpha = 10\%$  
   b) $\alpha = 5\%$  
   c) $\alpha = 1\%$  
   d) $\alpha = 0.1\%$  
   e) none of these

18. The 92% confidence interval for a population mean $\mu$ is $(13, 19)$. In a hypothesis test of $H_0 : \mu = 17$ versus $H_a : \mu \neq 17$ at the $\alpha = 8\%$ significance level, we would
   a) accept $H_0$  
   b) reject $H_0$  
   c) accept $H_a$  
   d) impossible to say  
   e) none of these

Use the following scenario for problems 19 through 21. *For a biology project, you measure the weight in grams and the tail length in millimeters (mm) of a group of mice. You then use this data to find the least-squares regression line for predicting tail length from weight: $\hat{y} = 20 + 3x$*

19. What tail length would you predict for a mouse that weighs 50 grams?
   a) 20 mm  
   b) 10 mm  
   c) 150 mm  
   d) 170 mm  
   e) none of these

20. How much on average does tail length increase for each additional gram of weight?
   a) 3 mm  
   b) 20 mm  
   c) 23 mm  
   d) 1/3 mm  
   e) none of these

21. The correlation coefficient between weight and tail length is $r = 0.80$. What fraction of the variation in tail length is explained by the least squares regression onto weight?
   a) 0.80  
   b) 0.64  
   c) 0.20  
   d) 0.36  
   e) none of these

Use the following scenario for questions 22 through 24. *Below is a histogram of light bulb lifetimes from a simple random sample of 30 bulbs taken from an assembly line. Note that the vertical axis is given in counts (i.e. number of bulbs). Assume that no data point fell exactly on a class (i.e. bin) boundary.*

22. Circle all words below that describe the shape of the distribution of light bulb lifetimes
   unimodal  bimodal  uniform  symmetric  left-skewed  right-skewed

23. What percent of light bulbs in the sample lasted less than 1100 hours?
   Answer:____________________

24. Which of the following classes contains the median light bulb lifetime?
   a) 1000 to 1100  
   b) 1100 to 1200  
   c) 1200 to 1300  
   d) 1300 to 1400  
   e) none of these
25. Below are three scatterplots, A, B, C. Rank them in order of least correlated to most correlated (i.e. from lowest correlation coefficient to highest).

Answer:____________________

Use the following scenario for problems 26 and 27. At right you see the histograms of four random samples, A, B, C, and D. They are drawn using the same scales on both the horizontal and vertical axis, so they can be directly compared against each other.

26. Which of the four samples (A, B, C, or D) has the largest sample mean?

Answer:____________________

27. Which of the four samples (A, B, C, or D) has the largest sample standard deviation?

Answer:____________________

28. In a χ² test for association, the two-way table has 4 rows and 3 columns. The degrees of freedom for the test statistic are

a) 3  b) 6  c) 7  d) 12  e) none of these

29. Your company makes detectors that monitor the quality of the air in deep underground mines. The detector contains a computer chip that is programmed to do hypothesis tests based on the concentrations of various impurities in the air. The null hypothesis is that the air quality is good. Which of the following would represent a Type II error?

a) declaring the air quality to be bad when in fact it is OK
b) declaring the air quality to be OK when in fact it is bad
c) computing an incorrect P-value for the test
d) failing to reject the null hypothesis
e) none of these