1. Based on data collected from the Salk field trial, the National Foundation of Infantile Paralysis would like to be able to say whether or not the polio vaccination was effective. At this point, which one of the following are they most concerned with? Choose one of the following.

(a) Data production.

(b) Data Description.

(c) Estimation.

(d) Test of significance.

2. Here is a small part of a data set that describes Major League Baseball players as of opening day of the 2007 season.

<table>
<thead>
<tr>
<th>Player</th>
<th>Team</th>
<th>Position</th>
<th>Age</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taveras, Willy</td>
<td>Rockies</td>
<td>Outfield</td>
<td>25</td>
<td>402</td>
</tr>
<tr>
<td>Beckett, Josh</td>
<td>Red Sox</td>
<td>Pitcher</td>
<td>27</td>
<td>6,667</td>
</tr>
<tr>
<td>Sabathia, C.C.</td>
<td>Indians</td>
<td>Pitcher</td>
<td>26</td>
<td>8,750</td>
</tr>
<tr>
<td>Rodriguez, Alex</td>
<td>Yankees</td>
<td>Third Base</td>
<td>31</td>
<td>22,709</td>
</tr>
</tbody>
</table>
3. Among persons aged 15 to 24 in the U.S., the leading causes of death and number of deaths in 2008 were: accidents, 14020; homicide, 5285; suicide, 4297; cancer, 1659; heart disease, 1059; congenital defects, 466.

(a) The variable “cause of death” is: (choose all that apply)

(i) qualitative

(ii) categorical

(iii) quantitative

(iv) numerical

(v) continuous

(vi) discrete

(b) The result “number of deaths” is: (choose all that apply)

(i) qualitative

(ii) categorical

(iii) quantitative

(iv) numerical

(v) continuous

(vi) discrete

4. What is the difference between a census and a sample survey? Choose one of the following.

(a) A census is expensive while sample surveys are relatively inexpensive to conduct.

(b) A census can study many variables whereas a sample survey can only study one or two.

(c) A census attempts to include the entire population in the survey whereas a sample survey only studies some of the members of the population.
(d) A census can only be conducted by a government whereas a sample survey can be conducted by anyone.

5. If a statistical study systematically favors certain outcomes, then it is called _________.  
Choose one of the following.

(a) A voluntary response sample

(b) A convenience sample

(c) A simple random sample

(d) Biased

6. Suppose a packaging inspector decides to inspect a sample from a crate of eggs for freshness. Each crate has 5 trays of 30 eggs each stacked on top of each other. The inspector decides to examine only the top row. Why is this sample biased? Choose one of the following.

(a) Each egg does not have an equal chance of being chosen in the sample.

(b) The top row may not be representative of the entire crate of eggs for freshness.

(c) The inspector was convenience sampling which is typically biased.

(d) All of the above.

7. A manufacturer of rubber wishes to evaluate certain characteristics of its product. It is known that their bales of synthetic rubber are stored on 300 pallets with a total of 15 bales on each pallet. They select a sample for inspection by first picking five pallets at random; then eight bales of rubber are randomly selected from each pallet. Does this scheme provide a simple random sample of bales? Yes or no. If yes, explain how this scheme satisfies the conditions necessary for simple random sampling. If no, specify which condition of simple random sampling is violated by this scheme.
8. In a survey of 1,500 UW students regarding the prevalence of binge drinking, 960 students (64%) said they participated in binge drinking on at least a monthly basis. Nationwide, the average proportion of college students who binge drink is 60%. Which of the following is the observed value for the statistic?

(a) 60%
(b) 957
(c) 64%
(d) 1500

9. Los Angeles has about four times as many registered voters as San Diego. A simple random sample of registered voters is taken in each city, to estimate the proportion who will vote for school bonds. Other things being equal, a sample of 1,000 taken in Los Angeles will have a margin of error that is

(a) four times as large
(b) twice as large
(c) as large

as the margin of error of a sample of 1,000 taken in San Diego. Choose one of the options.

10. The Margin of Error only covers (choose one)

(a) Non sampling errors
(b) Random sampling errors
(c) Undercoverage
(d) Nonresponse
11. A sample of households in a community is selected at random from the telephone directory. In this community, 4% of households have no telephone, 10% have only cell phones and another 25% have unlisted telephone numbers. This sample will certainly suffer from: (choose one)

(a) nonresponse.
(b) undercoverage.
(c) false responses.

12. Your friend is studying the perspective of U.W. students on whether or not they voted in the most recent election. She emailed out an Internet survey to 1,000 UW students, randomly generated from a list of UW students (you can assume the randomization was done correctly). However, there were several errors she was not aware of in conducting her study. Which of the following errors could apply to her study? Choose all that apply.

(a) Sampling error
(b) Non response error
(c) Convenience error
(d) Response error

13. Because the placebo effect is strong, designed experiments should be: (choose one)

(a) An observational study
(b) Randomized
(c) Double-blinded
(d) Biased

14. Use the following situation to answer parts a, b and c. The Nurses’ Health Study has interviewed a sample of more than 100,000 female registered nurses every two years since
1976. The study finds that “light to moderate drinkers had a significantly lower risk of death” than either nondrinkers or heavy drinkers.

(a) The Nurses’ Health study is: (choose one)

(i) a designed experiment.

(ii) an observational study.

(iii) can’t tell without more information.

(b) What is the explanatory variable? (choose one)

(i) Drinking habits of each nurse.

(ii) Female registered nurses.

(iii) Risk of death.

(c) What is the response variable? (choose one)

(i) Risk of death.

(ii) A yes/no variable for each nurse indicating whether she lived or died over the course of the study.

(iii) Number of deaths in the different categories of drinking habits.

15. The death rates of surgical patients differ for operations in which different anesthetics are used. An observational study found these death rates for four anesthetics.

<table>
<thead>
<tr>
<th>Anesthetic</th>
<th>Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halothane</td>
<td>1.7%</td>
</tr>
<tr>
<td>Pentothal</td>
<td>1.7%</td>
</tr>
<tr>
<td>Cyclopropane</td>
<td>3.4%</td>
</tr>
<tr>
<td>Ether</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

A statistician argues that this is not good evidence that cyclopropane is more dangerous than the other anesthetics since the risks inherent in different surgical procedures confounds the association. Which two facts listed below would have to hold simultaneously for risks of surgery to confound the relationship between choice of anesthetic and the death rate?
(a) The death rates for the anesthetics will reverse direction if risks inherent in surgery are controlled for.

(b) Doctors are more likely to prescribe cyclopropane for patients undergoing high risk surgery.

(c) The numbers of high risk surgeries are few.

(d) Patients undergoing high risk surgery are more likely to die.

16. A medical experiment compares an anti-depressant medicine with a placebo for relief of chronic headaches. There are 36 headache patients available to serve as subjects. To randomly choose 18 patients to receive the medicine, you would: (choose all that apply)

(a) Assign labels 00 to 35 and use random digits to choose 18.

(b) Assign labels 00 to 17 because only 18 need to be chosen.

(c) Put the 36 names in a hat, thoroughly mix and then pick 18.

(d) Arrange the names alphabetically and pick the first 18 names.

17. Graduate students in a statistics department recorded hours spent consulting on 21 different projects.

\[1\ 1\ 2\ 2\ 3\ 5\ 5\ 8\ 10\ 15\ 15\ 18\ 20\ 24\ 30\ 30\ 30\ 48\ 50\ 148\]

Just by looking at the data, describe the appearance of a graph of the data.

(a) Symmetric.

(b) Skewed to the right.

(c) Skewed to the left.

(d) Skewed in the middle.
18. On a test, 10 students scores were 76, 77, 81, 83, 88, 91, 91, 92, 94, and 95. Which of the following statements are true? Choose all that apply.

(a) The mean is higher than the median.
(b) The median is higher than the mean.
(c) The mode is higher than the mean.
(d) The mean and mode are equal.

19. A newspaper article compares median net worth for black, Hispanic, and white households, as reported by the Pew Hispanic Center: The medians in 2002 were approximately $6,000 for black households, $8,000 for Hispanic households, and $89,000 for white households. Four statistics students were asked to explain why researchers would report medians instead of means. Who gives the best answer?

(a) **Student 1:** Probably because medians are easier to compute.
(b) **Student 2:** I think they reported medians to make the results seem more dramatic, and make people aware of how discrimination against minorities can result in lower pay.
(c) **Student 3:** If they wanted results to seem more dramatic, they would have used means. The reason they used medians was so they wouldn’t be inflated by high outliers.
(d) **Student 4:** In case they don’t have the exact values for the households that they used, as long as the middle one is correct, the median will be accurate. Reporting the median is the best way to handle these data values.

20. The correlation between I.Q. scores and school G.P.A. is $r=0.634$. Which of the following is true? Choose all that apply.

(a) For every additional I.Q. point, a students GPA increases 0.634 on average.
(b) For every additional GPA point, a students I.Q. score increases 0.634 on average.
(c) A higher I.Q is associated with a higher GPA on average.
(d) I.Q. and GPA do not exhibit regression towards the mean.

21. (Use the situation described here to answer parts a, b and c.) The following figure shows histograms for two quantitative variables $X$ and $Y$, as well as a scatter plot of $X$ and $Y$.

(a) By looking at these plots, guesstimate the averages and standard deviations of $X$ and $Y$ and the correlation $r$ between $X$ and $Y$. Choose one of the following.

(i) Average of $X$ = 1, S.D. of $X$ = 1, Average of $Y$ = 10, S.D. of $Y$ = 10, $r$ = 0.7
(ii) Average of $X$ = 1, S.D. of $X$ = 1, Average of $Y$ = 5, S.D. of $Y$ = 4, $r$ = -0.7
(iii) Average of $X$ = 1, S.D. of $X$ = 2, Average of $Y$ = 5, S.D. of $Y$ = 10, $r$ = -0.7
(iv) Average of $X$ = 1, S.D. of $X$ = 2, Average of $Y$ = 1, S.D. of $Y$ = 4, $r$ = -0.1

(b) Using your chosen summary statistics from the previous question, calculate the least squares regression estimate for the average of $Y$ when $X$ is 3. Choose one of the following.

(i) Predicted average $Y$ is “-6.2”.
(ii) Predicted average $Y$ is “0”.
(iii) Predicted average $Y$ is “2.2”.

9
(iv) Predicted average $Y$ is “-0.6”.

c) What is the typical size of the error in your prediction from the previous problem? Choose one of the following?

(i) 2.8
(ii) 7.1
(iii) 0.7
(iv) 1.4

22. Suppose that 26% of Americans think the driving age should be raised to 18. An opinion poll interviews 600 randomly chosen Americans and records the sample proportion who feel that the driving age should be raised to 18. Due to sampling error, the statistic will vary from sample to sample if the poll is repeated. The sampling distribution is approximately Normal with mean 0.26 and standard deviation 0.02. What proportion of samples will indicate that more than 30% of Americans think the driving age should be raised? (choose one)

(a) 2.5%
(b) 16%
(c) 5%
(d) 26%

23. Four statistics students are discussing the answer to the following question: In Seattle, it rains on about 30% of the days in an average year. Does it make sense to report a probability of 0.3 (or 30%) for the chance of rain on any given day?. Choose the answer that offers the best reasoning.

(a) Student 1: I don’t think it makes sense to report a daily probability of rain of 0.3 because what actually happens and the probability of it happening aren’t directly related.
(b) **Student 2**: Besides, people are not interested in the forecast for the long run. Common people need a weather forecast every day so they will adapt to the weather.

(c) **Student 3**: I don’t think you can apply the probability of rain calculated on a yearly basis to individual days. I mean, days belong to different seasons and surely that has to affect the daily chances of rain.

(d) **Student 4**: They don’t report a probability of 30% every day because with every day that it does rain, it lowers the chances that will rain again. If it rains one day, then there is less than a 30% chance of rain for the next day.

24. Which of the following are valid interpretations of the Law of Large Numbers. Choose all that apply.

(a) If a random phenomenon with numerical outcomes is repeated many times independently, the mean of the actually observed outcomes approaches the expected value.

(b) It says how many trials are needed to guarantee an average outcome close to the expected value.

(c) It is valid for any sampling methods.

(d) When we toss a coin many times, the number of heads is getting close to half the number of tosses by the law of large numbers.

25. (Use the following situation to answer parts a, b, c, d, e). Suppose you were using an eight-sided number die that was rigged to have one side occur more than the others (not equally likely). The probability model of the trick die is:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.5</td>
<td>0.2</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>?</td>
</tr>
</tbody>
</table>

(a) What is the probability of rolling an “8”? Choose one.
(b) For the trick die in the previous problem, what is the probability of observing an odd number? Choose one.

(i) 0.5
(ii) 0.65
(iii) 0.7
(iv) Can’t tell from the information given.

(c) For the same trick die, what is the probability of observing an even number? Choose one.

(i) 0.35
(ii) 0.3
(iii) 0.5
(iv) Can’t tell from the information given.

(d) For the same trick die, what is the expected value for the number on the up-face of the die? Choose one.

(i) 0.5
(ii) 1
(iii) 2.55
(iv) Can’t tell from the information given.

(e) If this trick die is rolled 6000 times, then the average of the number of spots actually observed on each of the 6000 rolls should be exactly ___________. Choose one.
26. A new poll shows that 64% of Washingtonians support a temporary sales tax hike to mitigate cuts to education and social services. The Seattle Times report is based on a simple random sample of 408 Washington state voters and has a margin of error of 5%. Assume that the margin of error is calculated using a 95% confidence level. Which one of the following interpretations is correct?

(a) We are 95% confident that 64% of Washingtonian voters support a temporary sales hike.

(b) We are 95% confident that about 59% to 69% of Washingtonian voters support a temporary sales hike.

(c) We are 95% confident that about 59% to 69% of the sampled voters support a temporary sales hike.

(d) None of the above.

27. A poll conducted by the Siena College Research Institute in 2005 found that “81% of the adults surveyed would vote for a female president”. What other information about the experiment would be needed to construct a valid 95% confidence interval for the proportion of all adults in the population who would vote for a female president? Choose all that apply.

(a) How many people were sampled?

(b) How many women were polled?

(c) How many people in the population support a women president?
(d) How were the people sampled?

28. (Use the following scenario to answer parts a, b, c.) A few senior students at Ballard High want to estimate what proportion of the students in their school own an iPod. They gather an S.R.S of 500 students. They find that 187 of the students (37.4%) in the sample own an iPod.

(a) What is the population of interest for this poll? Choose one.

   (i) The senior class of Ballard High
   (ii) The students of Ballard High
   (iii) The students who participated in the survey
   (iv) None of the above.

(b) What is most likely true about the population proportion \( p \) in this setting? Choose one.

   (i) Exactly 37.4 %
   (ii) Approximately 37.4 %
   (iii) Not 37.4 %
   (iv) 187

(c) An approximate 95% confidence interval for the population proportion \( p \) is \( \ldots \). Choose one.

   (i) \( 0.374 \pm 2 \sqrt{\frac{0.374 \times (1-0.374)}{500}} \).
   (ii) \( 0.374 \pm 2 \sqrt{\frac{0.374 \times (1-0.374)}{187}} \).
   (iii) \( 37.4\% \pm 2 \sqrt{\frac{0.374 \times (1-0.374)}{500}} \).
   (iv) Cannot be calculated.
29. (Use the following scenario to answer parts a, b, c, d.) What proportion of adults in the United States always votes in the presidential elections? An article in American Demographics reports this percentage as 67%. To test this claim, a random sample of 800 adults was taken, and 512 of them (64%) said that they always vote in presidential elections.

(a) What are the null and alternative hypotheses? Choose one.

(i) \( H_0 : p = 0.67 \) vs \( H_a : p > 0.67 \)

(ii) \( H_0 : \hat{p} = 0.67 \) vs \( H_a : \hat{p} \neq 0.67 \)

(iii) \( H_0 : p = 0.67 \) vs \( H_a : p \neq 0.67 \)

(iv) \( H_0 : p = 0.64 \) vs \( H_a : p \neq 0.64 \)

(b) What is the standard deviation of the sampling distribution under the null hypothesis? Choose one.

(i) 0.017

(ii) 0.033

(iii) 0.470

(iv) 0.021

(c) What is the standard score for our observed \( \hat{p} \) and the P-value corresponding to the test of significance? Choose one.

(i) Standard score = -1.76 and P-value is 0.0801.

(ii) Standard score = -1.76 and P-value is 0.9199.

(iii) Standard score = 0.64 and P-value is 0.5157

(iv) Standard score = -1.76 and P-value is 0.0400

(d) Can this sample result be explained by chance? Suppose an investigator is using a significance level \( \alpha = 0.05 \). Which one of the following conclusions would he pick?
(i) The P-value suggests that the proportion of adults who vote is less than 67\% , but it is not very convincing evidence.

(ii) The P-value suggests that the proportion of adults who vote is less than 67\%, and it is very convincing evidence.