

Names: _____

MATH 232 · Introduction to Statistics

Spring 2017

Instructor: E.M. Kiley

Week 2: Cartoon Guide Questions

Please read pages 89–97, and answer the following questions together with your group mates. Please submit only one answer sheet per group, and clearly indicate the names of all group members at the top.

Question 1. What is funny about the title of the bedtime story our statistician friend is reading on p. 89? Do you know a quote or title that sounds similar to this?

Question 2. For now (until we learn about standard deviation), you can ignore the $\frac{1}{\sqrt{n}}$ mentioned on p. 91.

Question 3. On p. 93, the author brings up an issue surrounding the survey technique of randomly dialing telephone numbers. Last March, the Democratic primary election held in the state of Michigan resulted in one of the worst failures of polling prediction at its time. On the morning of the election, all of the nation’s leading pollsters and analysts, including Nate Silver¹, predicted that Hillary Clinton would beat Bernie Sanders by huge margins, ranging from 21 to 37 percentage points; in the actual contest that followed, Sanders would actually defeat Clinton by 1.5 percent of voters.

It is evident that the pollsters’ data samples were not representative of the population of Democratic voters in Michigan. The beginning of one possible explanation is that in Michigan, the law permitted political polling to be conducted only via landline telephones. Why might this have skewed the data in Clinton’s favor?

Question 4. The text on p. 94 says that making a sampling frame (that is, a list of every unit in the population from which to draw a random sample) might be “costly, controversial, or even impossible”. Give another scenario (in addition to the one about lakes mentioned in the book) in which making a list of every unit in the population might pose one or more of those issues.

¹Silver was a hobby statistician made famous in 2008 by his web site fivethirtyeight.com, which used statistical weighting methods to accurately predict the results of 49 out of 50 states in the general presidential election, and whose forecasts for **all** senate races that year were also accurate. He became interested in statistics because of a childhood love for baseball; it turned out to be very profitable for him when the New York Times subsequently hired him as a columnist, and later when ESPN purchased rights to his blog. He now has a team of employees working permanently at fivethirtyeight.com, which was one of the few outlets to ask what would happen to the predicted outcome of the 2016 presidential election if polls were off by certain small margins; in the end, their site gave the highest probability of the outcome that actually occurred.

Question 5. On p. 95, the author illustrates some strata of pickles (e.g., garlic dill, pickled peppers, etc.). Choose one of the other two examples from the top of p. 90 (voting populations or manufactured goods) and list some possible strata that could be useful in constructing a stratified sample.

Question 6. For cluster sampling, what do you hope is true about the clusters that end up in the sample, in order for the sample to be representative of the population?

Question 7. Which technique do you think is more expensive for a statistician to carry out: stratified sampling, or cluster sampling? Which costs do you need to account for?

Question 8. Do you feel that there is a significant difference between systematic sampling and simple random sampling?

Question 9. What is the problem with the data mentioned in the example of Shere Hite's questionnaire (an opportunity sample)? What makes the data biased?

Question 10. Can you think of a more timely example of opportunity sampling that might result from a careless use of modern technology?