Math 1011 Quiz 2 (100 points total, 10% of grade) 7 August, 2014

Name: Solutions

1. (15 points) Terminology

(a) What are the differences and similarities between a parameter and a statistic? In practice, which do you know and which do you want to know? What are they used for? (five points) A parameter is a number that describes the population while a statistic is computed from the sample. You know the statistic but want to know the parameter. The statistic is used to produce an estimate for the parameter

~~(b) What are the differences and similarities between quota sampling and multistage cluster sampling? Which is better? Why? (five points)~~

(c) Ian and John are curious how many people in their high schools play sports. Ian goes to a large school of 6000 students and John goes to a smaller school of only 2000. Each of them take a sample of size 200 and ask those students if they play sports. Since John talked to $1/10$ of his whole school and Ian only talked to $1/30$ of his school John concludes his results should be more precise. Ian does not agree. Who is right and why? (five points)

Ian is right. As long as the population is much larger than the sample (which it is at both schools) it does not affect the precision at all

2. (20 points) Standard Error

(a) Karen is doing a survey to predict the upcoming student council election. There are 2000 students that will vote and she asked 150 of them who they would vote for. She computed the SE of her data and sent it off to the school newspaper to be published (this school was exceptionally interested in statistics). Anyway, the next day the realizes that she forgot to multiply by the correction factor! She had computed the SE by modeling her sample as a random draw with replacement. Should she rush to the school newspaper and shout "stop the presses!" or should she not worry about it? Why? (five points)

Don't worry about it! Since the sample is a small part of the population the correction factor is close to 1.

(b) Larry stands in front of McDonald's and asks every third person who passes what country they were born in. He is interested in what percentage of his town was born in the US. Of the 100 people he asks 90 were born in the US and 10 were not. He computes the SE to be 3. He concludes that a 95% confidence interval for the percentage of people in his town born in the US is $90\% \pm 6\%$. Do you agree? Why or why not? (five points)

I do not agree. That formula only works for a simple random sample

(c) Micky takes a sample of 100 students from a large school and sees that 40 of them were interested in the World Cup. He (correctly) computes a 95% confidence interval for the percentage of the school interested in the World Cup. He then says "there is a 95% chance that the percentage of the school interested in the World Cup lies in this interval." Do you agree with what he said? Why or why not? Compute the 95% confidence interval. (ten points)

I do not agree. The percent of the school interested in the world cup is a fixed parameter. It does not change. He should say "There is a 95% chance that an interval created in this way will contain the percentage of students interested in the World Cup."

40% = center

$$\frac{\sqrt{100} \times \sqrt{\frac{4}{10} \times \frac{6}{10}}}{100} = \frac{2\sqrt{6}}{100} = 2\sqrt{6}\%$$

$$40\% \pm 4\sqrt{6}\% \approx 40\% \pm 9.8\%$$

3. (20 points) Collecting samples.

Nathan wants to take a sample of 100 people from his high school. He knows in advance that his school is exactly 50% female and 50% male. To take his sample he has himself and 9 of his friends each find 5 girls and 5 boys to interview. Critique this method of collecting a sample. (eight points)

Not a probability method so not good! He is very likely to get people similar to him and his friends since they have discretion about who they choose.

(b) How do quota sampling and weighting try to solve the same problem? What is the problem they are solving? How are they different and which is better? Why? (twelve points)

The problem is samples that are not representative of the population. Quota sampling (as above) is when interviewers have specific quotas they have to fill when interviewing BUT they can choose anyone they'd like within those quotas. With weighting you start by collecting a sample with a probability method & then give the responces of people different "weights" to counteract chance error. Weighting is better because it uses a probability method and there is no discretion by the interviewer.

4. (20 points) Averages

(a) Suppose you are interested in the average income of the people in your town. You take a simple random sample of 400 people. The sum of their incomes is \$4,000,000 and the SD is \$20,000. Compute a 95% confidence interval for the average income of the town. (ten points)

$SE_{avg} = \frac{\sqrt{400} \times 20{,}000}{400} = 1{,}000$

$avg = \frac{4{,}000{,}000}{400} = 10{,}000$

\$10,000 ± \$2,000

(b) In what part of this computation did you make an estimate? (five points)

I estimated the SD of the "box" (population) by the SD of the sample.

(c) Without doing all of those computations again, how can you find what the SE of the average would be if you instead took a sample of 1,600 people? (five points)

$1{,}600 = 4 \times 400$

Therefore new $SE = \frac{\text{old } SE}{2} = \frac{1000}{2} =$ \$500

5. (25 points) Computation

(a) Olivia takes a simple random sample of 400 people from her school to ask what grade they have in their math class. The sample has an average of 80 and a SD of 10. Compute a 95% confidence interval for the average math grade in this school. (ten points)

$$SE = \frac{\sqrt{400} \times 10}{400} = \frac{10}{20} = 1/2.$$

$$80 \pm 1$$

(b) But wait! Olivia notices that the sample she had of math grades is not even close to the normal curve. Can you justify the methods you used above to compute the confidence interval? (five points)

Yes! Even if the "box" is not normal the sum of draws is (for a large # of draws). This is the Central Limit Theorem. Since the sum is normal so is the average.

(c) Peter is curious about what percentage of his college was home schooled at some point in their life. He takes a simple random sample of 100 students and asks them. Out of the 100 exactly 10 say that they have been home schooled. Compute a 95% confidence interval for the percentage of students that are home schooled. (ten points)

$$SE = \frac{\sqrt{100} \times \sqrt{\frac{1}{10} \times \frac{9}{10}}}{100} = \frac{1}{10}\sqrt{\frac{1}{100}} = \frac{3}{100} = 3\%$$

$$10\% \pm 6\%$$

(extra space to work if you need it)