

Stat 1040 Recitation 9 Solutions

1. A child knows a game in which she has a 1 in 3 chance of winning, a 1 in 3 chance of losing and a 1 in 3 chance of a tie. Suppose the child gets a prize if she wins more than 35% of the time.

(a) Is she more likely to win the prize if she plays 100 times or 1000 times? Explain briefly using the law of averages.

We expect her to win around  $33\frac{1}{3}\%$  of the time. She gets a prize if the % error is large and this is more likely for 100 because % error decreases as the number of times increases.

(b) Suppose the child plays the game 100 times. Find the chance that she wins more than 35% of the time.

The % of times she wins is like the % of 1's in 100 draws from the box

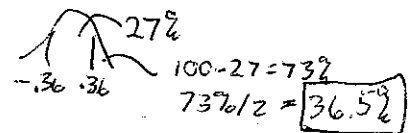
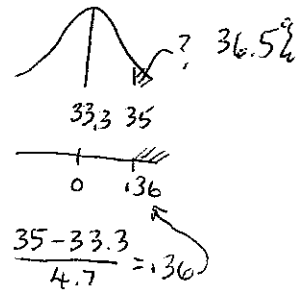
2 0 1     $ave_{box} = .333$   
 $SD_{box} = .47$

$$EV_{sum} = 100(.33) = 33.3$$

$$EV_{\%} = 33\%$$

$$SE_{sum} = \sqrt{100} (.47) = 4.7$$

$$SE_{\%} = \frac{4.7}{100} \times 100\% = 4.7\%$$



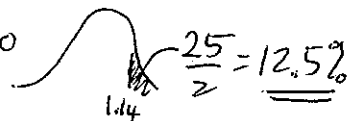
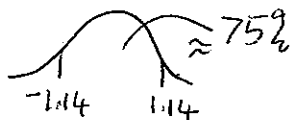
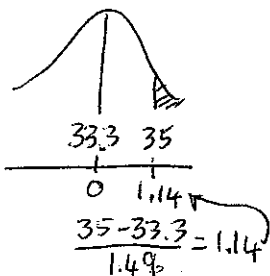
(c) Suppose the child plays the game 1000 times. Find the chance that she wins more than 35% of the time.

As above but now 1000 draws.

$$EV_{\%} = 33.3\%$$

$$SE_{sum} = \sqrt{1000} (.47) = 31.6(.47) = 14.9$$

$$SE_{\%} = \frac{14.9}{1000} \times 100\% = 1.49\%$$



(d) Do your answers for (b) and (c) confirm what you said in (a)?

Absolutely!

2. In a board game, a player rolls two dice and adds up the number of spots on the two dice to determine how far to move. Consider the following two box models.

A. The number of spots is like the sum of two draws from a box with 6 tickets labeled 1, 2, 3, 4, 5, 6.

B. The number of spots is like one draw from a box with 11 tickets labeled 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Which box is correct, or are they the same? Explain.

A is correct. For box B, it looks as though getting a total of 6 has the same chance as a total of 2 and this is not true (6 is much more likely than 2 from the chart in chapter 14).

3. A gambler plays a game of chance. In each game, one of the following 3 things happens:

- Win \$1
- Lose \$1
- No change

(a) The probabilities for these three things are 0.3, 0.4, and 0.3 respectively. Each game is independent of the rest. If the gambler plays this game 400 times, find a box model for the net amount they win.

The net amount they win is like the sum of 400 draws from the box:

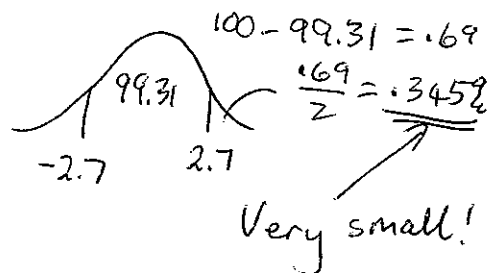
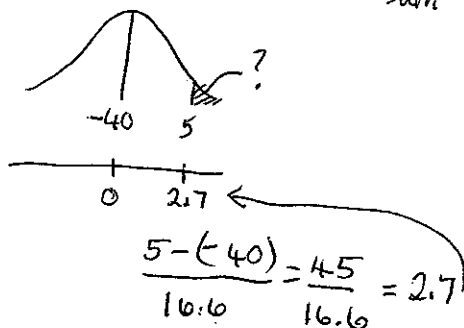
3 [1] 4 [-1] 3 [0]

(b) Find the chance the gambler will win more than \$5.

3 [1] 4 [-1] 3 [0]     $ave_{box} = -.1$   
 $SD_{box} = .83$

$$EV_{sum} = 400(-.1) = -40$$

$$SE_{sum} = \sqrt{400}(.83) = 16.6$$



4. An elementary education teacher teaches in two different elementary schools. The first school has 500 children and the second school has 1000 children. She is going to randomly sample a total of 30 students. If she wants to get equal accuracy for the two schools, should she sample 10 children from the first school and 20 children from the second school? If so, why? If not, what should she do? Explain.

To get equal accuracy, she should sample 15 from each school because (assuming everything else is comparable) accuracy depends only on sample size.

5. All other things being equal, which is more accurate, taking a simple random sample of 300 eggs from a farm that produces 6,000 eggs daily, or taking a simple random sample of 400 eggs from a farm that produces 12,000 eggs daily? Explain.

The sample of size 400 is more accurate because for simple random sampling, larger samples are more accurate (population size does not matter as long as the populations are large).

6. A newspaper web site has a poll on gun control. In one hour, they had 1200 votes - 850 "for" gun control and the rest "against" gun control.

- (a) Explain why we would not expect these 1200 votes to be like a simple random sample of the general population.

The general population do not all use the web and even those who use the web may not read that website or take the poll.

- (b) Would they even be a simple random sample of people who read that newspaper site? Explain. No.

People who feel strongly are more likely to respond.  
People may respond several times.

Some people are less likely than others to do polls.

- (c) Give a plausible source of bias in the survey.

Perhaps people who read that web site are more likely to be "for" gun control than the general population.  
Perhaps Republicans are more likely "for" + more likely to use the web.

7. A health inspector randomly chooses 10 one-dozen cartons of "large AA" eggs from a store (so she gets 120 eggs). Is this a simple random sample of eggs from the store? If not, what sort of a sample is it? Explain.

No. It's a cluster sample because she either gets all the eggs in a carton or none of the eggs in that carton.

8. The National Collegiate Athletic Association requires colleges to report the graduation rates of their athletes. At USU, 62% of all 157 student athletes who entered between Fall 1990 and Fall 1993 graduated within 6 years of starting. True or false, and explain:

An approximate 95% confidence interval for the percentage of all USU athletes who graduate within 6 years of starting runs from 54.3% to 69.7%.

False, because we have no sample. We have the whole population of USU athletes for that time period.

9. Suppose researchers selected a simple random sample of 1,200 U.S. taxpayers and found that in 2004 these 1,200 received an average tax refund of \$2,052, with a standard deviation of \$431. According to the IRS, the average refund for all U.S. taxpayers that year was \$2,063.

- a. What does it mean when we say that the researchers selected a "simple random sample" of 1,200 U.S. taxpayers?

Taxpayers were selected at random like drawing tickets at random from a box, without replacement.

- b. Using the researchers' results, find a 90 percent confidence interval for the average refund for all U.S. taxpayers in 2004.

sample ave = 2052

ave box = ??

SD box = ??  $\approx$  431

$$SE_{sum} \approx \sqrt{1200} (431) = 14930$$

$$SE_{ave} \approx \frac{14930}{1200} = 12.44$$

$$CI: \$2052 \pm (1.65)(12.44) = \$2052 \pm \$20 \quad \$2032 - \$2062$$

- c. True or False, and explain briefly: The sizes of the tax returns of 90% of all U.S. taxpayers in 2004 are in this 90 percent confidence interval.

False - the confidence interval is saying we're 90% confident the average of all taxpayers is between \$2032 and \$2062. Only a small percentage of individual taxpayers will be in that interval.

10. The U.S. Bureau of Labor Statistics regularly collects information on the labor market. According to the bureau, workers employed in manufacturing industries earned an average \$577 per week in May 1999. Assume that this average is based on a simple random sample of 1600 workers selected from manufacturing industries and that the standard deviation of weekly earnings in this sample is \$80.

- (a) Find a 99% confidence interval for the average weekly earnings of all U.S. workers employed in manufacturing industries in May 1999.

$\begin{array}{c} \text{99\%} \\ \text{---} \\ -2.6 \quad 2.6 \end{array}$   
 2.6 is the multiplier

$\text{mean}_{\text{box}} = ??$   
 $\text{SD}_{\text{box}} = ?? \times 80$

$$SE_{\text{sum}} = \sqrt{1600} (80) = 3200$$

$$SE_{\text{ave}} = \frac{3200}{1600} = 2$$

$$CI: 577 \pm 2.6(2) = \$577 \pm 5.2.$$

- (b) We know that income does not follow the normal curve. Since this is the case, is the confidence interval you found in part a) still valid? Explain.

Yes, it's still valid because the sample size is large.

11. A student takes a simple random sample of 500 USU students and finds that 374 of them picked up a copy of the "Statesman" in the last week. Find an approximate 95% confidence interval for the percentage of all USU students who picked up a copy of the "Statesman" in the last week.

374 out of 500 is  $\frac{374}{500} \times 100\% = 74.8\%$

bootstrap:  $\frac{25 \mid 0}{75 \mid 1}$        $\text{mean}_{\text{box}} = .75$   
 $\text{SD}_{\text{box}} = .433$

$$SE_{\text{sum}} = \sqrt{500} (.433) = 9.68$$

$$SE_{\%} = \frac{9.68}{500} \times 100\% = 1.94\%$$

CI is  $74.8 \pm 2(1.94)\%$

$$74.8 \pm 3.9\%$$

$$\underline{\underline{70.9\% \text{ to } 78.7\%}}$$