Karle Page 1

Question 1 (10 points)
In tossing a fair coin, follow the steps to find the standard error (SE) that makes the chances equally likely in the two scenarios.

a) 10 +/- 2 heads in 20 tosses is about as likely as 490 +/- ? heads in 980 tosses.

Step 1: Compare the number of tosses in the 2 cases. 980 is 49 times more than 20. (1 point)

Step 2: Sum or percent? (**Circle your answer** – 1 point)
Multiply or divide? (**Circle your answer** – 1 point) by $\sqrt{49}=7$ (fill in the blank with a number – 1 point)

Step 3: Your new SE is 14. (1 point)

$2 \times 7 = 14$

b) 50% +/- 9% heads in 8 tosses is about as likely as 50% +/- ? heads in 648 tosses.

Step 1: Compare the number of tosses in the 2 cases. 648 is 81 times more than 8. (1 point)

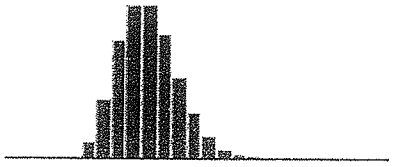
Step 2: Sum or percent? (**Circle your answer** – 1 point)
Multiply or divide? (**Circle your answer** – 1 point) by $\sqrt{81}=9$ (fill in the blank with a number – 1 point)

Step 3: Your new SE is 1. (1 point)

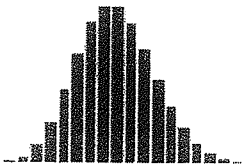
$9/9 = 1$

Question 2 (3 points)
The 3 histograms below (in scrambled order) are the probability histograms for the sum of 40, 100, and 300 random draws with replacement from a box that has 49 tickets marked "0" and 1 ticket marked "1".

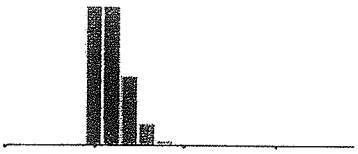
Histogram A



Histogram B



Histogram C

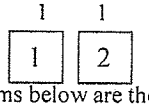


Fill in the blanks below to match the histograms with the correct number of draws.

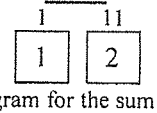
a) Histogram C is the probability histogram for 40 draws from the box.
b) Histogram A is the probability histogram for 100 draws from the box.
c) Histogram B is the probability histogram for 300 draws from the box.

Question 3 (3 points)
30 draws are made at random with replacement from each of the following boxes:
The number above each ticket represents how many of the tickets there are. For example, box C has 1 ticket marked "1" and 25 tickets marked "2."

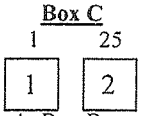
Box A



Box B

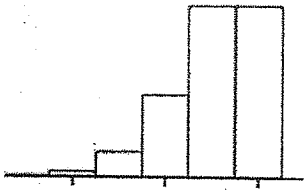


Box C

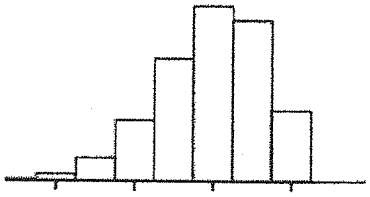


The 3 histograms below are the probability histogram for the sum of 30 draws from Box A, Box B, and Box C.
Which histogram is which? *Fill in the blanks above each Box with A, B, or C.*

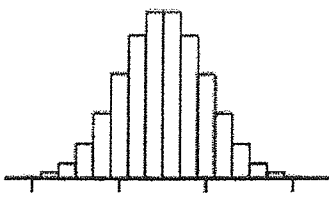
Box C



Box B



Box A



Karle Page 2

Question 4 (17 points)
Fill in the first blank with the number of draws, the second with either "with" or "without" and the third with the letter corresponding to the appropriate box model. Choose from the box models below. *Use each box model exactly once.*

Box A

1	0

Box B

2	-1

Box C

1	2	3	4	5	6

Box D

1	0

a) A die is rolled 3 times and the sum of the spots is counted.
This corresponds to drawing 3 times with replacement from Box C.

b) A die is rolled 10 times and the number of 2's is counted.
This corresponds to drawing 10 times with replacement from Box A.

c) A fair coin is tossed 20 times. If the coin lands on heads you win $2. If it lands on tails you lose $1.
This corresponds to drawing 20 times with replacement from Box B.

d) A true/false test has 50 questions. Suppose you randomly guess on all 50 questions and your score is the number of correct answers you get.
This corresponds to drawing 50 times with replacement from Box D.

2 pts e) What is the SD of Box B? 1.5

$SD = |2-(-1)|\sqrt{\frac{1}{2}\times\frac{1}{2}} = 1.5$

f) What is the SD of Box D? (hint: no work necessary) 0.5

g) The smallest the SD of a 0-1 box can be is 0 and the largest the SD of a 0-1 box can be is 0.5.

Question 5 (15 points)
25 draws are made at random with replacement from the box containing these 5 tickets:

0	2	3	4	6

a) The smallest the sum of the 25 draws could possibly be is 0 and the largest is 150.
(Fill in the 2 blanks above with the correct numbers.) 2 pts 2 pts

b) What is the EV (expected value) of the sum of the 25 draws? (Show work, circle answer.) 2 pts

$EV_{sum} = n \times \text{ave of box} = 25 \times 3 = (75)$

c) What is the SE (standard error) of the sum of the 25 draws? Use the fact that the SD of the box is 2. (Show work, circle answer.) 2 pts

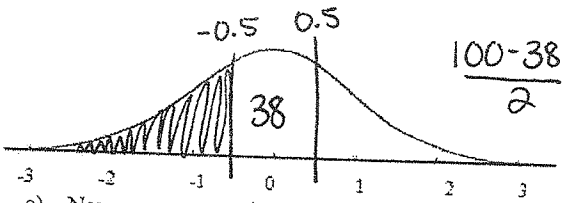
$SE_{sum} = \sqrt{n} \times \text{SD of box} = \sqrt{25} \times 2 = (10)$

d) Use the normal approximation and your answers from (b) and (c) above to figure out the chance that the sum of the 25 draws will be less than 70.

i. First calculate the z-score (1 point) $Z = \frac{\text{val}-EV}{SE} = \frac{70-75}{10} = -0.5$

ii. Next mark the z-score on the curve and shade correctly (1 point)

iii. Write the chance that the sum of the 25 draws will be less than 70 in the box below. (1 point)



$\frac{100-38}{2} = 31$

Chance = 31 %
(Round to nearest whole number.)

e) Now suppose you draw at random with replacement from the same box above, but this time you're only interested in counting how many 2's you get. What is the EV and SE for the number of 2's in 25 draws? (Hint: draw a new box!)

i. EV of the number of 2's in 25 draws = 5 2 pts

$EV_{sum} = n \times \text{ave of box} = 25 \times \frac{1}{5} = 5$

ii. SE of the number of 2's in 25 draws = 2 2 pts
(Hint: use your new box to find the SE)

$SE_{sum} = \sqrt{n} \times SD \text{ of box} = \sqrt{25} \times 0.4 = 2$

1	4
1	0

5 tickets

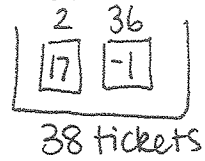
$\text{ave} = \frac{1}{5}$

SD = 0.4

Karle Page 3

Question 6 (19 points)
A gambler plays roulette 100 times betting \$1 on two numbers (7 and 11) each time. If the ball lands on either 7 or 11, the gambler wins \$17. If the ball lands on any of the other 36 numbers the gambler loses \$1. The roulette wheel has 38 total slots numbered 1-36, 0, 00.

a) Which is the appropriate box model? Circle one: (2 points)
- i) The box has 38 tickets: 1 marked "7" and 1 marked "11" and the rest marked "0".
- ii) The box has 38 tickets: one of each marked 1, 2, 3,…. 36, 0, and 00.
- iii) The box has 38 tickets: 2 marked "17" and 36 marked "0".
- (iv) The box has 38 tickets: 2 marked "17" and 36 marked "-1".



b) How many draws from the box? 100 (1 point)

c) The draws are made…. *Circle one*: (i) with replacement ii) without replacement (1 point)

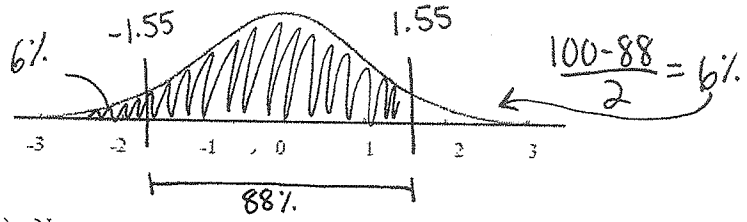
d) What is the average of the box? Write your answer as a fraction. Show work and circle answer. (2 points)

$$ave = \frac{2(17) + 36(-1)}{38} = \frac{-2}{38}$$

e) What is the SD of the box? Show work and circle answer. Round your answer to 2 decimal places. (2 points)

$$SD = |17-(-1)|\sqrt{\frac{2}{38} \times \frac{36}{38}} = 4.02$$

f) Use the normal approximation and the fact that the EV= \$ -5 and the SE= \$40 (approximately) to figure out the chance that the gambler will win **less than \$57** in 100 plays?
- *i.* First calculate the z-score (1 point) $Z = \frac{val - EV}{SE} = \frac{57-(-5)}{40} = 1.55$
- *ii.* Next mark the z-score on the curve and shade correctly (1 point)
- *iii.* Write the chance that the gambler will win **less than \$57** in 100 plays in the box below. (1 point)

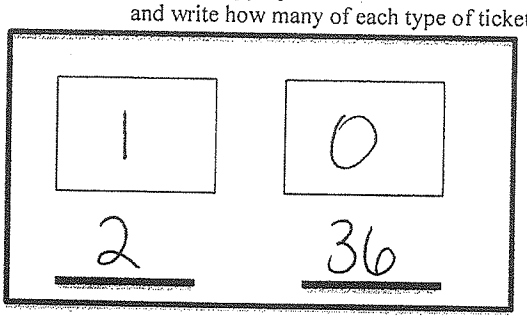


Chance = 94 %
(Round to nearest whole number.)

88 + 6 = 94

g) Now suppose we are only interested in how many times we'd expect the gambler to win playing 100 times (instead of how many dollars we'd expect him to win).

Draw the appropriate box model in the box below. Label the 2 rectangular tickets inside the box with the correct values and write how many of each type of ticket in the blank below each ticket. (4 points)



i) What is the average of the new box? **Show work. Leave answer as a fraction.** (1 point)

$$ave = \frac{1(2) + 0(36)}{38} = \frac{2}{38}$$

ii) What is the SD of the new box? **Show work. Round answer to 2 decimals.** (1 point)

$$SD = |1-0|\sqrt{\frac{2}{38} \times \frac{36}{38}} = 0.22$$

h) Now suppose we play 400 times instead of 100 times.

Part 1) How will this affect the Standard Error of the sum?
- i) It will increase (be multiplied) by 4.
- (ii) It will increase (be multiplied) by 2.
- iii) It will decrease (be divided) by 4.
- iv) It will decrease (be divided) by 2.

Part 2) How will this affect the Standard Error of the average?
- ii) It will increase (be multiplied) by 4.
- iii) It will increase (be multiplied) by 2.
- iv) It will decrease (be divided) by 4.
- (v) It will decrease (be divided) by 2.

Karle Page 4

Question 7 (4 points)
A recent NBC News Poll asked a random sample of 500 adults nationwide the following question: "Do you support legalizing marijuana?" At the same time CNN posted the same question on their website as a "Quick Vote" questions where anyone who wants to can cast their vote. Here are the results of both surveys.

	No	Yes	Sample Size
NBC News Poll	52%	48%	500
CNN Quick Vote	90%	10%	8,360

a) For which poll is it appropriate to calculate the margin of error?
Choose one:
i) NBC News Poll because NBC says on their website that they are "fair and balanced."
(ii) NBC News Poll because the people were randomly drawn from all adults nation-wide.
iii) CNN Quick Vote survey because it has more people.

b) As you can see, the results of the two polls are quite different. Which survey gives a better estimate of the percentage of all US adults who would say that they support legalizing marijuana?
Choose one:
i) NBC News Poll because NBC says on their website that they are "fair and balanced."
(ii) NBC News Poll because the people were randomly drawn from all adults nation-wide.
iii) CNN Quick Vote survey because it has more people.

Question 8 (4 points)
a) Suppose a government survey organization is planning to take a simple random sample of people in each state in order to estimate the percentage of college graduates in that state. The want to take a random sample of people in New York (population = 25 million) and a random sample in Montana (population = 1 million). All other things being equal, to achieve the same level of accuracy in both polls, the number of people you'd have to poll in New York is about _____ the number of people you'd have to poll in the Montana. (2 points)

i) 25 times smaller than **ii)** 5 times smaller than (iii) the same as iv) 5 times larger than v) 25 times larger than.

b) About how many people would you have to poll in New York to get a 95% Confidence Interval with a Margin of Error of **10%?** **(Assume the SD of the population is close to 0.4)** (1 point)

(i) 64 ii) 100 iii) 625 iv) 1111 v) 2500 vi) 1 million

$$n = \left(\frac{200 \times 0.4}{10}\right)^2 = 64$$

c) About how many people would you have to poll in the Pennsylvania to get a 95% Confidence Interval with a Margin of Error of **3%? (Now let's assume the SD of the population is close to 0.5)** (1 point)

i) 64 ii) 100 iii) 625 (iv) 1111 v) 2500 vi) 10,000

$$n = \left(\frac{200 \times 0.5}{3}\right)^2 = 1111.11$$

Question 9 (5 points)
For the following 5 questions, circle whether you think the statement is True or False.

a) Parameters are numerical facts about the sample.
i. True (ii.) False

b) Inferences are generalizations about the population that come from the sample.
(i.) True **ii.** False

c) If we do not have a random sample, we cannot calculate standard errors or confidence intervals.
(i.) True **ii.** False

d) Samples are not only representative of the population that they are drawn from, but they are also representative of subgroups of that population.
i. True (ii.) False

e) The Central Limit Theorem says that the probability histogram of all possible sums (or averages, or percentages) of draws from any box will get closer and closer to the normal curve as the number of draws increases.
(i.) True **ii.** False

Karle Page 5

Question 10 (8 points)
A Harris Poll asked a random sample of 1,000 nationwide male adults the following question: "Are you afraid of clowns?" 10% of the people in the sample answered "YES". The SE of the sample percent is about 1%.

a) An approximate 95% confidence interval for the percentage of all American men who are afraid of clowns is:
Show work below and write your final answer in the blanks below with the smaller number on the first line and larger number on the second line. (2 points)

(8 , 12)

95% CI = sample % ± 2(SE%)
= 10 ± 2(1)

b) An approximate 77% confidence interval for the percentage of all American men who are afraid of clowns is:
Show work below and write your final answer in the blanks below with the smaller number on the first line and larger number on the second line. (2 points)

(8.8 , 11.2)

77% CI = sample % ± 1.2(SE%)
= 10 ± 1.2(1)

c) Which confidence interval will be narrower? (2 points)
i) The 95% Confidence Interval
(ii) The 77% Confidence Interval
iii) Impossible to tell which will be narrower
iv) They will be the same.

d) Suppose 100 pollsters each randomly sampled 1,000 male adults nationwide asking whether they were afraid of clowns. All 100 pollsters computed 80% confidence intervals to estimate the percentage of all US male adults who are afraid of clowns. Would all 100 intervals correctly include the true population percent? (2 points)
Choose one:
i) Yes, all 100 would include the true population percent assuming no errors were made.
(ii) No, only about 80 of them would include the true population percent.
iii) No, we don't know how many would include the true population percent.

Question 11 (12 points)
Suppose I wanted to study students' attention span during my 80 minute STAT 100 lectures. To do this, I chose a random sample of 64 students out of the 1100 students enrolled in my class to follow in detail. The average number of minutes the 64 students reported not listening during the 80 minute lecture was 20 minutes with an SD of 10 minutes.

a) Which most closely resembles the relevant box model? (2 points)
Choose one:
i) The box has 1100 tickets marked with "1"s and "0"s.
ii) The box has 64 tickets marked with "1"s and "0"s.
(iii) The box has 1100 tickets, marked with numbers ranging from 0 to 80. The exact average and SD of the box are unknown and estimated from our sample.
iv) The box has 1100 tickets with an average of 20 and an SD of 10.

b) How many draws from the box? 64 (1 point)

c) The draws are made...**Circle one**: (1 point) i) with replacement (ii) without replacement

d) The best estimate for the average number of minutes all 1100 students would report not listening is 20. (2 points)

e) What is the SE of the sample average? Show work. Circle answer. (2 points)
$SE_{ave} = SD/\sqrt{n} = 10/\sqrt{64} =$ (1.25 or 10/8)

f) Is it possible to compute a 95% confidence interval for the average number of minutes all 1100 students report not listening from the information given? (2 points)
Choose one:
(i) Yes, since our sample was random.
ii) No, because we aren't given the SD of the sample.
iii) No because a sample is not representative of the population it was drawn from.

g) Circle the statement(s) that are true. (2 points)
i) The expected value for the average of not listening in class for all of my male STAT 100 students is 20 minutes.
ii) The expected value for the average of not listening in class for all U of I students is 20 minutes.
(iii) The expected value for the average of not listening in class for all of my STAT 100 students is 20 minutes.