Question 1 (4 pts.)
For the following questions circle the more likely possibility. If they're equally likely circle both:

a. (i) Playing roulette 100 times, betting $1 on red each time and coming out ahead unusual
(ii) Playing roulette 900 times, betting $1 on red each time and coming out ahead.

b. (i) Tossing a fair coin 10 times and getting exactly 50% heads. unusual
(ii) Tossing a fair coin 100 times and getting exactly 50% heads.

c. (i) Tossing a fair coin 100 times and getting between 45 and 55 heads SE=5 , Z=-1,1 Ch=68%
(ii) Tossing a fair coin 400 times and getting between 190 and 210 heads. SE=10, Z=-1,1 Ch=68%

d. (i) Tossing a fair coin 100 times and getting between 40%-60% heads.
(ii) Tossing a fair coin 1000 times and getting 40%-60% heads. usual

Question 2 pertains to the following situation: (14 pts.)
100 draws are made at random with replacement from a box that has 5 tickets: one each of 1,3,4,5,7.

a) What is the smallest possible the sum of the draws can be? 100 What is the largest? 700

b) What is the EV for the sum of the draws? 100*4

c) What is the SE for the sum of the draws? (The SD of the box is 2) SE= $\sqrt{n}$ *SD= 10*2=20

d) Use the normal curve to estimate the chance that the sum of the draws is between 395 and 405? **Show work. Circle answer.** (Give your answer to 2 decimal places.)

Z= (405-400)/20=0.25 Z= (395-400)/20=-0.25
Area between -.25 and .25 on the normal curve = 19.74%

e) How many "7"s would you expect in 100 draws? 1/5 * 100 = 20

f) What is the SE for the number of "7"s in 100 draws? **Show work. Circle answer.** (Hint: draw a new box)

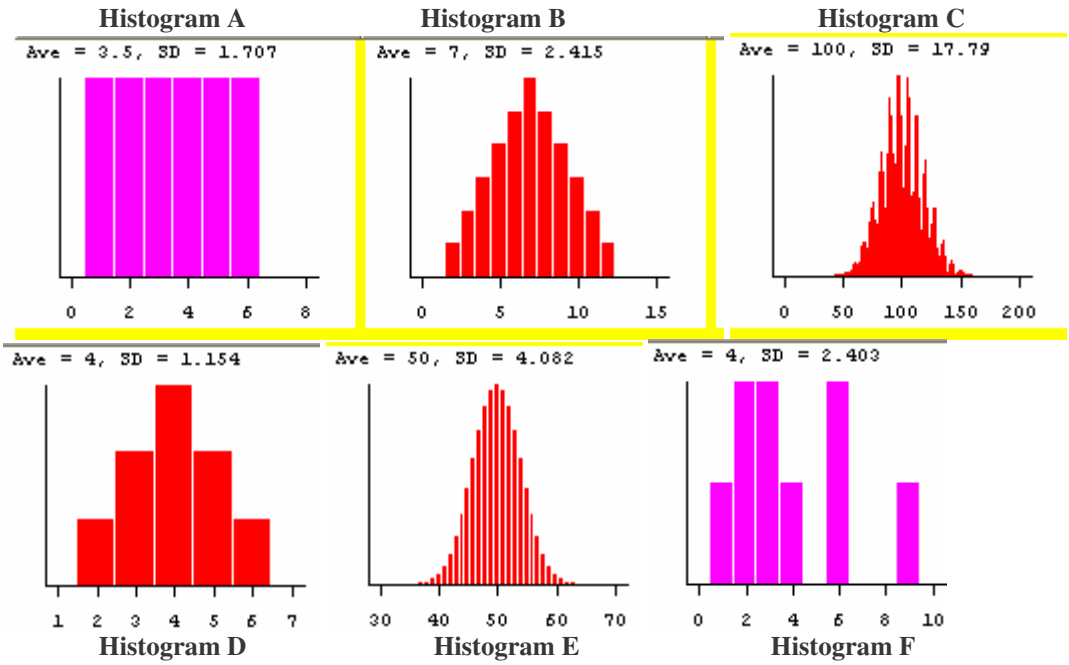
New Box 1 0 0 0 0 SD = $\sqrt{(1/5)(4/5)}$ = 2/5 SE = $\sqrt{100}$ * 2/5 = 4

g) The EV for the number of odd numbers in 100 draws from the box above is 80 and the SE for the number of odd numbers drawn is 4. Use the normal curve to estimate the chance of drawing between 75-85 odd numbers in 100 draws? **Show work. Circle answer.** Give your answer to 2 decimal places.

Z = (85-80)/4 = 1.25
Area between -1.25 and 1.25 on the normal curve = 78.87%

Question 3 pertains to the 3 boxes and 3 histograms below: (8 pts.)

Box 1 contains 6 tickets: 1, 2, 3, 4, 5, 6 Box 2 contains 3 tickets: 1, 2, 3 Box 3 contains 3 tickets: 1, 2, 9



Circle A, B,C,D,E or F below to make the statements true. For (a) and (b) also circle the number of the correct Box. Each histogram is circled exactly once.

a) Histogram **A** B C D E F is the histogram for the **contents** of Box **1** 2 3.
b) Histogram A B C D E **F** is the probability histogram for the **product** of 2 draws from Box 1 **2** 3
c) Histogram A B C D **E** F is the probability histogram for the **sum** of **25** draws from Box 2.
d) Histogram A B **C** D E F is the probability histogram for the **sum** of **25** draws from Box 3.
e) Histogram A B C **D** E F is the probability histogram for the **sum** of **2** draws from Box 2.
f) Histogram **A** **B** C D E F is the probability histogram for the **sum** of **2** draws from Box 1.

Question 4 (2 pts.)
A simple random sample eliminates what type of bias?
a) Only selection bias
b) Only non-response bias
c) Only response bias (response bias deals with how the questions are phrased).
d) All the above

Question 5 (6 pts.)
The CNN website conducts a public opinion poll daily called QuickVote. Any Internet user can go to www.cnn.com and cast their vote. Last Thursday the QuickVote question was: "Did you think Osama bin Laden would be captured or killed by now?" 202,227 people voted, 65% voted YES and 35% voted NO.

a) What type of sample is this? i) Probability Sample ii) Quota Sample iii) Neither

b) The main problem with this sample is i) Sample Size ii) Bias in the wording iii) Selection Bias

c) What is the SE for the percentage of "YES"s?

i) $\sqrt{202{,}227} * \sqrt{(.65)(.35)}$ ii) $\frac{\sqrt{(.65)(.35)}}{\sqrt{202{,}227}} * 100\ \%$ iii) SE is not valid here b/c it's not a scientific survey

Question 6 pertains to the following: (9 pts.)

The U of I enrolls about 38,000 students, of whom 26% are graduate students. A simple random sample of 400 students is drawn.

a) The EV for the % of the sample that is graduate students is 26 % and the SE for the % is 2.19%
Show work for computing the SE. Round your answer to 2 decimal places.

$\sqrt{(.26)(.74)/400} * 100\% = 2.19\%$

b) Part (a) corresponds to drawing 400 times at random without replacement from which box?
Fill in the first blank with a number and the second with either "with" or "without". Circle the correct box model below.

i) The box has 400 tickets, 26% marked "1" and 74% marked "0"
ii) The box has 38,000 tickets, 26% marked "1" and 74% marked "0"
iii) The box has 38,000 tickets marked with "1"s and "0"s. The exact percentages are unknown but are estimated from the sample.

c) Suppose 50% of the 38,000 students are female. The SE of the sample % that is female= 2.5% .Use the normal curve to figure the chance that between 49% and 51% of the sample of 400 is female? Show work and draw a picture.

Z= (51-50)/2.5 = .4 Area between .4 and -.4 on the normal curve is 31%

Question 7 (6 pts.)
The average family income of the 10,000 undergraduates enrolled at a certain private university is $100,000 with an SD of $60,000. A simple random sample of 400 is drawn from the population of 10,000.

a) What is the EV for the average income of the sample? $100,000 same as for the population
b) What is the SE for the average income of the sample? Show work. Circle answer.

SE = SD/$\sqrt{n}$ = \$60,000/$\sqrt{400}$ =\$3000

c) The chance that the sample **average** is between $40,000 and $160,000 is closest to

i) 68% ii) 95% iii) 100% iv) Not possible to calculate since income does not follow the normal curve

It doesn't matter whether the incomes themselves follow the normal curve because with enough draws the AVERAGE incomes will follow normal curve (Central Limit Theorem)
Z= (Value – EV)/ SE = (160,000-100,000)/3000 = 20

Question 8 (4 pts.)

A Harris Poll asked a random sample of 1,015 adults nationwide the following question: "Are you very afraid of snakes?" 36% of the people in the sample answered "YES".

a) The SE of the sample % is 1.5%. An approximate **68%** confidence interval for the percentage of all American adults who are afraid of snakes is closest to:

i) (34.5%-37.5%) ii) (30%-36%) iii) (32%-40%) iv) (35.985%-36.015%)

b) If the researcher increased the sample size to 4, 060 the length of the confidence interval would

i) be multiplied by 4 ii) double iii) be divided by 4 iv) be divided by 2 v) not possible to calculate

Question 9 pertains to the following situation: (8 pts.)

A CNN/USA Today/Gallup Poll conducted Aug 25-26 2003 found that 42% of adult Americans answered "Yes" to the question: "Would you vote for Arnold Schwarzenegger if he were running for governor in your state? "
The poll was based on telephone interviews with 1,009 people nationwide. The sample of telephone exchanges was randomly selected from a complete list of all telephone exchanges in the country. Then random digits were added to form a complete telephone number. Within each household one adult was designated by a random procedure to be the respondent for the survey. Assume all Americans have telephones.

a) What type of sample is this?
i) Probability Sample ii) Quota Sample iii) Sample of Convenience iv) Self-selected sample

b) What most closely resembles the relevant box model? Circle one.
i) It has 1,009 tickets, 42% marked "1"and 58% marked"0".
ii) It has millions of tickets, with an average of 0.42, but the SD is unknown.
iii) It has millions of tickets marked with "0"s and "1"s. The exact percentages are unknown but are estimated from the sample.
iv) It has millions of tickets, 42% marked "1"and 58% marked"0".

c) Calculate the SE for the percentage of people in the sample who answered "YES". Show work, circle answer. (Round to 2 decimal places.)

$\sqrt{(.42)(.58)/1009} * 100\% = 1.55\%$

d) Suppose the poll was only taken in Illinois (instead of nationwide), how should CNN adjust the sample size to keep the same SE?
i) Significantly increase sample size ii) Significantly decrease sample size iii) Keep sample size about the same

Question 10 (8 pts.)
A Gallup Poll asked a random sample of 1572 people nationwide the following question : "Just thinking about your own situation, how much money per year would you need to make in order to consider yourself rich?"
The sample average was $120,000 and the sample SD was $158,600

a) What most closely resembles the relevant box model? Circle one.
i) It has 1572 tickets marked with "0"s and "1"s.
ii) It has about millions of tickets marked with "0"s and "1"s, but the exact percentage of each is unknown.
iii) It has millions of tickets. On each ticket is written a dollar amount. The exact average and SD are unknown but are estimated from the sample.
iv) It has 1572 tickets. The average of the tickets is $120,000 and the SD is $158,600

b) What is the SE of the sample average? (Show work, circle answer. Round your answer to the nearest dollar.)

$\$158,600/\sqrt{1572} = \$4000$

c) A **68%** confidence interval for the average amount of money all American adults think they'd need to make per year in order to consider themselves rich is closest to

i) $110,000-130,000 ii) $116,000-124,000 iii) Can't calculate since data doesn't follow normal curve.

d) If the study asked the 1572 people whether or not they think they'll ever be rich, the relevant box model would contain tickets with

i) Only “1”s and “0”s ii) Dollar amounts

Question 11 (14 pts.)
Fill in the first blank with the number of draws and the second blank with the word "with" or "without", then circle the appropriate box model.

a) A gambler plays roulette 100 times betting a $1 on the number "7" and "11" each time. If the ball lands on either "7" or "11" he wins $17, if it lands on any other number he loses $1. There are 38 numbers: 0,00,1,2,3,…,36. This corresponds to taking the sum of 100 draws with replacement from which of the following box models?

i) The box has 100 tickets, 2 marked "17" and the rest marked "-1"
ii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
iii) The box has 38 tickets, 18 marked "1", 18 marked "-1" and 2 marked "0"
iv) The box has 38 tickets, 1 marked "35" and 37 marked "-1"
v) The box has 38 tickets, 2 marked "17" and 36 marked "-1"

b) A multiple-choice test has 30 questions. Each question has 3 possible answers, only 1 of which is correct. Your score is computed as the number right minus half the number wrong. Suppose you guess at random on each question and your score is computed. This corresponds to taking the sum of 30 draws with replacement from which of the following box models?

i) The box has 30 tickets, half marked "1" and half marked "-1"
ii) The box has 3 tickets, 1 marked "1" and 2 marked "0"
iii) The box has 3 tickets, 1 marked "1", and 2 marked "-1/2".
iv) The box has 30 tickets, ten are marked "1" and twenty are marked "0".
v) The box has 3 tickets, 1 marked "1", and 2 marked "-1".

c) You roll a die 60 times and count the number of "4"s. This corresponds to taking the sum of 60 draws with replacement from which of the following box models?

i) The box has 6 tickets, 1 marked "4" and 5 marked "0".
ii) The box has 6 tickets, 1 marked "1" and 5 marked "0".
iii) The box has 6 tickets: one each of 1,2,3,4,5,6.
iv) The box has 60 tickets: 10 each of 1,2,3,4,5,6.

d) Two dice are tossed and the total number of spots is counted. Consider the following 3 box models:

- Box A-The box has 11 tickets: one each of 2,3,4,5,6,7,8,9,10,11,12
- Box B-The box has 6 tickets, one each of 1, 2,3,4,5,6
- Box C- The box has 36 tickets: 1 marked "2", 2 marked "3", 3 marked "4", 4 marked "5", 5 marked "6", 6 marked "7", 5 marked "8", 4 marked "9", 3 marked "10", 2 marked "11", 1 marked "12"

Which is the correct box model? (The number of draws may change depending on the box.) Circle one.

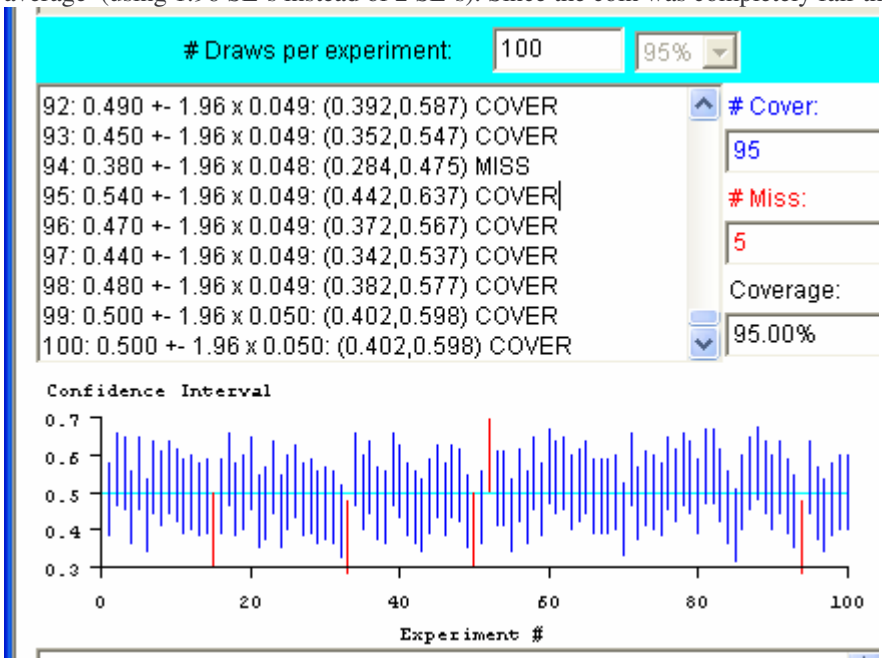
i) Only Box A
ii) Only Box B
iii) Only Box C
iv) All of them
v) Box A and B are correct. C is not.
vi) Box B and C are correct. A is not. Either 2 draws with replacement form Box B or 1 draw from Box C

Question 12 (4 pts.)
A little M&M Halloween treat packet contains 25 candies give or take 2 candies or so. (In other words the average number of candies per little packet is 25 and the SD is 2.) The stores sell the little packets in big bags. Each big bag contains exactly 36 little treat packets.

a)You would expect a big bag to contain 900 candies give or take 12 or so. (Fill in the first blank with the correct EV and the second with the correct SE) 25 *36 = 900 SE for sum = sqrt of n * SD = sqrt of 36 * 2 =12

b) About 20%, give or take 8% or so, of the M&M candies in each little packet are red. You would expect 20 % of the M&M candies in the big bag to be red give or take 1.33 % or so. (Fill in the first blank with the correct EV and the second with the correct SE, given to 2 decimal places.) Divide 8% by the sqrt of 36 (SE for % goes down by a factor of sqrt of n)

Question 13 pertains to the figure below copied from the Box Model Program for Confidence Intervals. (5 pts.)
The figure shows the results of a computer simulation: 100 imaginary experimenters each tossed the same coin 100 times to determine whether or not it was a fair coin. The coin really was fair, but none of them knew that. They each counted the number of heads they got and divided by 100 to get the sample average and then computed 95% confidence intervals for the true average (using 1.96 SE's instead of 2 SE's). Since the coin was completely fair the true average is 0.5.



a) How many of the 100 confidence intervals would you expect to miss the true average of 0.5? 5 because 90% confidence interval means you're 95% sure you cover the true average , so 5% of the time you miss the true average and 5% or 100 is 5.

b)

c) If the 100 experimenters each computed 90% confidence intervals (instead of 95% confidence intervals), how many would you expect to miss the true average of 0.5? 10 because 90% confidence interval means you're 90% sure you cover the true average , so 10% of the time you miss the true average and 10% of 100 is 10

c) If each experimenter only tossed the coin 10 times, instead of 100 would the length of each confidence interval change, and if so, how?

i) It wouldn't change
ii) It would increase by a factor of 10
iii) It would decrease by a factor of 10
iv) It would increase by a factor $\sqrt{10}$ because the sample is 10 times smaller so error will be sqrt of 10 times larger.
v) It would decrease by a factor $\sqrt{10}$

d) 5 of the 100 intervals in the figure above missed the true average of 0.5. If there were 1000 experimenters each tossing the coin 100 times and computing 95% confidence intervals, how many of the 1000 would you expect to miss?

i) 5 ii) 50 iii) $5 * \sqrt{10}$ iv) $5/\sqrt{10}$

Because a 95% confidence interval means you expect to miss 5% of the time and 5% of 1000 = 50

e) Experimenter #94 found a confidence interval of *0.38* +/- 1.96 x .048 How many heads did he get? 38
Because a 95% confidence interval = sample average +/- 1.96 SE's, so the sample average is 0.38= 38 heads out of 100.

Question 14 pertains to the following situation: (8 pts.)
Suppose the British Imperial Yard is sent to Paris for calibration against The Meter. Its length is measured 100 times. The average of the 100 measurements is 91.4402 cm, and the SD is 800 microns. (A micron is a millionth of a meter.) (You may assume the Gauss model with no bias.)

a) A single measurement is likely to be off by around _____microns i)80 ii) 8 iii) 800

Because the SD applies to individual measurements

b) The SE for the average of the measurements is _____microns. i)80 ii) 8 iii) 800

SE for the average = SD/sqrt of n = 800/ sqrt of 100 = 80

c) The number of tickets in the relevant box model is i) 91 ii)100 iii) Unspecified, but very large

d) The draws are made _____ replacement. i) With ii) Without

Because you can get the same measurement more than once.