Question 1

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

a) A gambler plays roulette 100 times betting a $1 on red each time. If the ball lands on red the gambler wins $1, if the ball lands on black or green the gambler loses $1. The roulette wheel has 18 red slots, 18 black slots and 2 green slots. **This corresponds to drawing _____times _____replacement from which of the following box models?**

Circle one:

i) The box has 100 tickets: 50 marked "1" and 50 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: 18 marked "1" and 20 marked "-1"
v) The box has 38 tickets: 18 marked "1" and 20 marked "0"

b) A multiple-choice test has 100 questions. Each question has 4 possible answers, only 1 of which is correct. Suppose you guess at random on each question and your score is computed as the total number you answer correctly minus ¼ the number you got wrong. **This corresponds to drawing _____times _____replacement from which of the following box models?**

Circle one:

i) The box has 4 tickets: 1 marked "1" and the 3 marked "0"
ii) The box has 2 tickets: 1 marked "1" and 1 marked "0"
iii) The box has 4 tickets: 1 marked "1", and 3 marked "-1/4".
iv) The box has 100 tickets: half are marked "1" and half are marked "0".
v) The box has 4 tickets: 1 marked "1", and 3 marked "-1".

c) A pair of dice are rolled and the total number of spots are counted. Which are the appropriate box models?

Circle exactly two: (Hint: one of the boxes would require drawing twice and the other would require only 1 draw.)

i) The box has 6 tickets, one each of 1,2,3,4,5,6
ii) The box has 11 tickets one each of 2,3,4,5,6,7,8,9,10,11,12
iii) The box has 6 tickets, 1 marked "1" and 5 marked "0"
iv) 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”

d) Consider this gambling game. You roll a pair of dice and if you get doubles (doubles means both dice show the same number of spots) you win $5. If you don’t get doubles you lose $1. You play this game 10 times. **This corresponds to drawing _____times _____replacement from which of the following box models?**

Circle one:

i) The box has 6 tickets, one each of 1,2,3,4,5,6
ii) The box has 11 tickets one each of 2,3,4,5,6,7,8,9,10,11,12
iii) The box has 6 tickets, 1 marked "5" and 5 marked "-1"
iv) The box has 10 tickets, half marked "5" and half marked "-1"

Questions 2 a-c pertain to tossing a fair coin and 2d pertains to guessing on every question of a true/false test- (4 pts.)

For each of the following circle the ***more*** likely possibility:

a) Getting exactly 5 heads in 10 tosses	***or***	Getting exactly 5000 heads in 10,000 tosses
b) Getting between 40%-60% heads in 100 tosses	***or***	Getting between 40%-60% heads in 10,000 tosses
c) Getting more than 3 heads in 10 tosses	***or***	Getting more than 30 heads in 100 tosses
d) Randomly guessing on every question of a 10 question true/false test and getting at least 60%	***or***	Randomly guessing on every question of a 100 question true/false test and getting at least 60%

correct.

correct.

Question 3 pertains to the following situation: (12 pts.)
400 draws are made at random with replacement from the box containing 4 tickets: 2 4 4 10
The average of the box is 5 and the SD of the box is 3.

a) The smallest the sum of the 400 draws could possibly be is _____and the largest the sum could be is _____
.

b) The expected value for the **sum** of the draws is .._____

b) The SE of the **sum** of the draw …_____

c) Use the normal curve to estimate the chance that the sum of the draws is greater than 4,060. Draw a picture, shade the correct area and circle answer.

d) What is the expected value for the **average** of the 400 draws?_____

e) What is the SE of the **average** of the 400 draws? _____

f) What is the expected value for the **number** of 10's drawn in 400 draws? _____

g) What is the SE of the **number** of 10's drawn in 400 draws? Show work. Circle answer and round to 2 decimal places. (Hint: draw a new box.)

Question 4 (6 pts.)
The MSNBC-TV website has a "Question of the Day" poll where anyone can cast a vote on various issues. One week in the middle of the Clinton-Lewinsky scandal, more than 200,000 people went to the website and voted on whether President Clinton should leave office. 73% of the 200,000 voted "Yes". That same week, an NBC News-Wall Street Journal poll asked the same question of 2,000 randomly selected adults nationwide and found that only 34 % voted "Yes".

a) Which poll gives a better estimate of what all US adults thought at the time about whether Clinton should leave office?

i) The MSNBC-TV because a survey of 200,000 people is more reliable than a survey of 2000 people in estimating what all 200 million Americans think
ii) The NBC News-Wall Street Journal poll is a better estimate because people who read the Wall St Journal are more informed than people who respond to polls online
iii) The NBC News-Wall Street Journal poll because the people were randomly selected.
iv) The two polls will have about the same degree of accuracy because the advantages and disadvantages of each will balance The advantage of large size is balanced by the disadvantage of selection bias for one poll while the advantage of random selection is balanced by the disadvantage of small size for the other.

b) What is the SE of the percentage of YES's in the Wall St Journal sample?
Circle one:

i) $\frac{\sqrt{.34*.66}}{\sqrt{2000}}$ **x100%** ii) $\sqrt{2000}\text{x}\sqrt{.34*.66}$ iii) $\frac{\sqrt{.34*.66}}{2000}$ **x100%** iv) Impossible to compute a SE for this sample.

c) If the number of people who voted online increased from 200,000 to 800,00 the accuracy of the poll would …
Circle one:
i) increase by a factor of 4
ii) decrease by a factor of 2
iii) increase by a factor of 2
iv) impossible to calculate the change in accuracy with this type of sample

Question 5 pertains to the following situation: (6 pts.)
In roulette, there are 38 numbers, 0,00,1,2,3,4,…36. Consider betting $1 on the 3 numbers 1,2, and 3.. If the ball lands on any of the 3 numbers, you win $11, but if the ball lands on any other number, you lose $1. The average in the corresponding box is $ -.0526, and the SD in the box is $3.24. Imagine playing this bet 100 times.

3a) What is the appropriate box model?

i) The box has 100 tickets: 3 marked “11”, 35 marked “-1” and the rest marked “0”.
ii) The box has 38 tickets: 3 marked "11" and 35 marked "-1"
iii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
iv) The box has 38 tickets: 1 marked “35” and 37 marked “-1”
v) The box has 38 tickets: 18 marked "1" and 20 marked "-1"

3b) The expected value of the sum of the draws is closest to

i) $5.26 ii) $ -5.26 iii) $32.40 iv) $-.324 v) $-.526

3c) The standard error of the sum of the draws is closest to

i) $5.26 ii) $ -5.26 iii) $32.40 iv) $-.324 v) $-.526

Question 6 pertains to the following situation: (14 pts.)
A recent Gallup poll asked a simple random sample of 900 adults nationwide how many pounds they'd like to lose.
The sample average **was 15 lbs.** with a **SD of 10 lbs.**

a) What most closely resembles the relevant box model?
Circle one.
i) It has 900 tickets marked with "0"s and "1"s.
ii) It has about millions of tickets marked with "0"s and "1"s..
iii) It has about 200 million of tickets. On each ticket is written a weight. The exact average and SD are unknown but are estimated from the sample.
iv) It has 900 tickets. The average of the tickets is 15 lbs. and the SD = 10 lbs.

b) The draws are made _____replacement. With Without

c) What is the SE of the sample average? (Show work, circle answer.)

d) Circle whether each of the statements below is true or false:

i) 15 +/- 0.67 lbs. is a 95% confidence interval for the average amount of weight that all American adults would like to lose.
True False

ii) 15 +/- 0.67 lbs. is a 95% confidence interval for the average amount of weight that all American males would like to lose.
True False

iii) It’s not valid to construct confidence intervals for the average amount of weight that all American adults because the sample data does not follow the normal curve.
True False

iv) It’s not valid to construct confidence intervals for the average amount of weight that all American adults because the draws are made without replacement. True False

e) An approximate 80% confidence interval for the population average is the sample average +/- _____* SE.
Circle the correct number to make the statement true.
i) 2.9 ii) 0.8 iii) 1.1 iv) 1.3 v) 1.6

f) Suppose the poll was taken just in Illinois (instead of nationwide) how should the pollsters adjust the sample size to keep the same SE? *Circle one*:

i) Significantly increase sample size ii) Significantly decrease sample size iii)Keep sample size about the same

Question 7 (3 pts.)

The Census Bureau is planning to take a simple random sample of 1000 in each state to estimate the percentage of the population with more than 12 years of education. Other things being equal, the accuracy to be expected in New York (population=20million) is _____the accuracy in Montana (population = 1/2 million).

i) quite a bit lower than ii) about the sameas iii) quite a bit higher than

Question 8 pertains to the 4 boxes and 5 histograms below: (10 pts.)

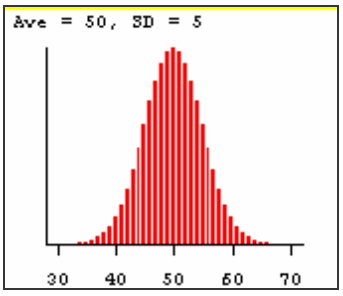
Box 1
0 1 2

Box 2
0 1

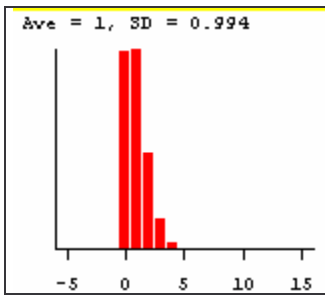
Box 3
9 0 's 1

Box 4
99 0 's 1

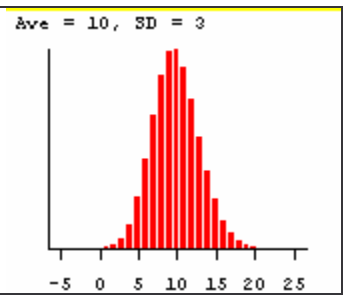
Histogram A

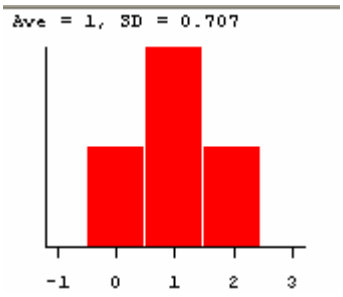


Histogram B

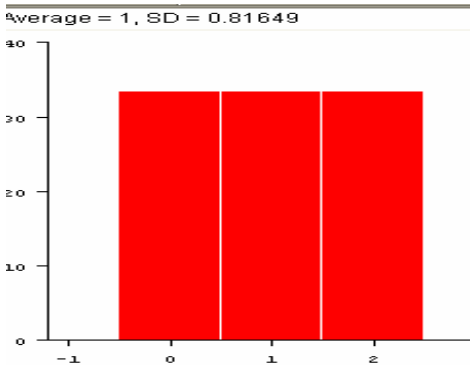


Histogram C





Histogram D



Histogram E

Fill in the blanks below to identify the correct histogram. Use each histogram exactly once. (10 pts)

a. The histogram for the **contents** of Box 1 is Histogram_____

b. The probability histogram for the **sum** of 100 draws from Box 4 is Histogram___.

c. The probability histogram for the **sum** of 100 draws from Box 2 is Histogram____.

d. The probability histogram for the **sum** of 100 draws from Box 3 is Histogram___

e. The probability histogram for the **sum** of 2 draws from Box 2 is Histogram___

Question 9 pertains to the following situation: (8 pts.)
According to the National Institute of Allergy and Infectious diseases, 20% of the United States adult population is infected with HSV-2, the virus that causes genital herpes. Assume this percent to be accurate.

a) What is the expected value for the **percent** of HSV-2 infected adults in a random sample of 100 U.S. adults?_____%

b) What's the chance that less than 10% of a random sample of 100 US adults would be infected with HSV-2?
Circle the closest answer. Show work for full credit (unless you answer iv*)*

i

i) Less than 1% ii) About 5% iii) About 20% iv) Not possible to calculate from the information given

c) What's the chance that more than 26% of a random sample of 100 US adult females would be infected with HSV-2? ?
Circle the closest answer. Show work for full credit (unless you answer iv*).*

i) About 50% ii) About 16% iii) About 26% iv) Not possible to calculate from the information given

d) Is it possible to compute a 95% confidence interval for the percent of all adults worldwide who are infected with HSV-2 from the information given?

i) Yes, it would be 20% +/- 2 * $\sqrt{\mathbf{.2*.8}}$ *100%
ii) No, because we're not given the SD of the sample
ii) No, because the data from the United States is not a worldwide random sample

Question10 (14 points)
A CBS news poll asked a simple random sample of 878 parents nationwide: **"Would you be willing to pay an additional $100 a year in taxes if the money were to be used to help set up an alternative school in your community?"**
60% answered "Yes" and 40% answered "No"

a) What most closely resembles the relevant box model? Circle one.
i) It has 878 tickets, 60% are marked "1" and 40% are marked "0"
ii) It has millions of tickets; on each ticket is written a dollar amount, but the exact average and SD are unknown.
iii) It has millions of tickets marked "0" and "1", but the exact percentage of each is unknown.

b) The draws are made _____replacement. *Circle one:* :i) With ii) Without

c) A 95% confidence interval for the percentage of all Americans parents who would answer "Yes" to the above question is closest to: *Circle one:*

i) (55%-65%) ii) (58.35-61.65%) iii) (56.7%-63.3%) iv) not enough information is given to determine

d) If the researcher increased his sample size by a factor of 9 (to n = 7902) then the length of the 95% confidence interval would…… *Circle one:*

i) increase by a factor of ii) increase by a factor of 9 iii) decrease by a factor of 3 iv) decrease by a factor of 9

e) Now suppose the poll had asked the random sample of 878 parents nationwide this question instead:

"How much would you be willing to pay in additional taxes each year to support an alternative school in your community?" And suppose the responses ranged from $0 to $1000 with an average of $100.

i) What most closely resembles the relevant box model? *Circle one*:
- It has 878 tickets marked "0" and "1", but the exact percentages of each are unknown.
- It has millions of tickets; on each ticket is written a number ranging from 0 to 1000. The average of the tickets is $100 but the SD is unknown.
- It has millions of tickets; on each ticket is written a dollar amount but the exact average and SD are unknown.

ii) A 95% confidence interval for the average amount all Americans parents would give to support an alternative school in their community is closest to: *Circle one:*
- ($90-$110)
- ($66.25-$133.74)
- ($83.13-$116.87)
- not enough information is given to calculate a confidence interval

Question12 pertains to the following situation: (10 pts.)
A Fox News poll conducted on October 25-26, 2005 asked a random sample of 900 adults nationwide the following question: "Do you personally believe in the existence of ghosts.?" I asked you the same question on the last Bonus Survey. Here's the results of both surveys:

	Believe	Don't Believe	Don't Know	Sample Size
Fox News random poll	34%	62%	4%	900
Bonus Survey 3	46%	30%	24%	494

a) As you can see, the results of the 2 polls are quite different. Which survey gives a better estimate of the percentage of all US adults who would say they believe in ghosts?
Circle one:
i) Fox News survey because it has more people
ii) The Bonus Survey because the responses are probably more honest
iii) The Fox News survey because the people were randomly drawn.

b) The SE of the percentage of people in the Fox News sample who answered "YES" is closest to…
Circle one:
i) $\frac{\sqrt{.34*.66}}{\sqrt{900}}*100\%$ ii) $\frac{\sqrt{.34*.62}}{\sqrt{900}}*100\%$
iii) Impossible to compute a SE for this sample because it can't be translated into a box model.
iv) Impossible to compute a SE for this sample because there are too many people who answered "Don't Know".

c) The SE of the percentage of people in the Bonus Survey sample who answered "YES" is closest to …
Circle one:
i) $\frac{\sqrt{.46*.30}}{\sqrt{494}}*100\%$ ii) $\frac{\sqrt{.46*.54}}{\sqrt{494}}*100\%$
iii) Impossible to compute a SE for this sample because it can't be translated into a box model.
iv)Impossible to compute a SE for this sample because there are too many people who answered "Don't Know".

d) Suppose another polling organization also randomly sampled 900 adults nationwide on October 25-26,2005 but this sample had 40% who answered that they believed in ghosts. Both polls construct 95% confidence intervals. Which of the following are true statements.
i) Only one of the confidence intervals could be right. (In other words, the true population percent could only be within one of the intervals.)
ii) Neither of the confidence intervals could be right
iii) It's possible that both confidence intervals are right, that is both could include the true population percentage.
v) If our calculations are correct, both of the confidence intervals will include the true population percent.

e) Suppose 100 pollsters all randomly sampled 900 adults nationwide and each computed **80%** confidence intervals and 95% confidence intervals. About how many of the 80% confidence intervals would miss the true population percentage? _____

About how many of the 95% confidence intervals would miss?_____