STUDY GUIDE FOR EXAM 3 covering Chapters 12-18

For each of the topics below look over the relevant problems in your lecture notes and homework.

Chapter 12: Box Models

What numbers should go on the tickets? How many of each type? With or without replacement?

Practice Problems:

Translate the following games of chance into box models by circling the appropriate box, specifying the number of draws and circling with or without replacement.

1) A 100 question multiple-choice test awards 1 points for each correct answer and subtracts 1/2 point for each incorrect answer. Each question has 3 choices only 1 of which is correct. Suppose you guess at random on each question.

a) Which is the appropriate box model?
Circle one:
- i) The box has 3 tickets: 1 marked "1" and 2 marked "-1/2"
- ii) The box has 2 tickets: 1 marked "1" and 1 marked "-1"
- iii) The box has 3 tickets: 1 marked "1", and 2 marked "0".
- iv) The box has 100 tickets, all marked either "1" or "-1/2", but the exact percentages of each type are unknown
- v) The box has 3 tickets: 1 marked "1" and 2 marked "-1"

b) How many draws from the box? __________

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

2) A gambler plays roulette 100 times betting $5 on red each time. If the ball lands on red the gambler wins $5, if the ball lands on black or green the gambler loses $5. The roulette wheel has 18 red slots, 18 black slots and 2 green slots.

a) Which is the appropriate box model?
Circle one:
- ii) The box has 100 tickets: 50 marked "5" and 50 marked "-5"
- iii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
- iv) The box has 38 tickets: 18 marked "5", 18 marked "-5" and 2 marked "0"
- v) The box has 38 tickets: 18 marked "1" and 20 marked "-1"
- vi) The box has 38 tickets: 18 marked "5" and 20 marked "-5"

b) How many draws from the box? __________

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

3) A fair die is rolled 60 times and the number of "3"'s is counted.

a) Which is the appropriate box model?

Circle one:

i) The box has 6 tickets, 1 marked "3" 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.

b) How many draws from the box? ___________

c) The draws are made ….

Circle one: **i)** with replacement **ii)** without replacement

4. 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	Box B	Box C	Box D
1 2 3 4 5 6	0 1	-1 1	0 0 0 1 1 0

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

b) A die is rolled 20 times and the total number of 4's and 5's are counted.
This corresponds to drawing _________times __________replacement from Box ______

c) A die is rolled 10 times and you win $1 if you roll an even number but lose $1 if you roll an odd number.
This corresponds to drawing _________times __________ replacement from Box ____

d) A coin is tossed 100 times and the number of heads is counted.
This corresponds to drawing _________times __________ replacement from Box ____

Chapter 13- The EV and SE of the Sum

- **Expected Value of the sum of n draws, made at random with replacement from a box:**

$$\mathbf{EV_{sum} = n \times \text{ave of the box}}$$

- **Standard Error of the sum of n draws, made at random with replacement from a box:**

$$\mathbf{SE_{sum} = \sqrt{n} \times \text{SD of the box}}$$

This is called the Square Root Law

- **Short –Cut Formula for computing the SD of a box with ONLY tickets marked "a" and "b":**

$$\mathbf{SD = |a-b| \sqrt{(\text{fraction of "a" tickets})(\text{fraction of "b" tickets})}}$$

Practice Problems:

1. **100** draws are made at random with replacement from the box containing 7 tickets:
$\boxed{-6}\ \boxed{-4}\ \boxed{-2}\ \boxed{0}\ \boxed{2}\ \boxed{4}\ \boxed{6}$
The SD of the box is 4.

a) The smallest the sum of the **100** draws could possibly be is _______and the largest the sum could be is________

b) The Expected Value for the **sum** of the **100** draws is ___________

c) The Standard Error of the **sum** of the **100** draws is________

2) A gambler plays roulette **100** times betting $1 on the numbers 1, 2 and 3 each time. If the ball lands on 1, 2 or 3 the gambler **wins $11**, if the ball lands on any of the other 35 numbers the gambler **loses $1**. The roulette wheel has 38 slots numbered 1-36, 0 and 00.

a) Which is the appropriate box model? ***Circle one:***
- i) The box has 38 tickets: 1 marked "1", 1 marked "2", 1 marked "3"and 35 marked "-1"
- ii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
- iii) The box has 38 tickets: 3 marked "11" and 35 marked "-1"
- iv) The box has 38 tickets: 3 marked "11" and 36 marked "0"

b) How many draws from the box? __________

c) What is the average of the box? Write your answer as a fraction.

d) What is the SD of the box? ***Show work. Circle answer. Round your answer to 2 decimal places.***

e) What is the Expected Value of the gambler's winnings? (Show work.)

f) What is the SE of the gambler's winnings? (Show work.)

e) Now suppose you're only interested in counting how many ***times*** the gambler would be expected to win if he played 100 times? What is the **EV** and the **SE** of the ***number*** of times he'd win if he played **100** times? ***(Hint: draw a new box with only 1's and 0's)***

i. EV of the ***number*** of $\boxed{1}$' s in **100** draws =__________ (2 pts.)

ii. SE of the ***number*** of $\boxed{1}$' s in **100** draws =___________ (2 pts.)
(Show work by computing the SD of the new box, then use it to calculate the SE for the sum of the 100 draws.)

Chapter 14—The EV and SE of the Average and Percent

- **EV of the average of n draws is just the average of the box.**
- **EV of the percent of 1's in n draws from a 0-1 box is just the percent of 1's in the box.**
- **SE of the average of n draws = SD of the box/ $\sqrt{n}$**
- **SE of the percent of 1's in n draws = SD of the box/ $\sqrt{n}$** (x 100%)

- **Law of Averages: Compare SE of the sum to SE of the average and percent.**

The SE for sums GROWS (is multiplied by $\sqrt{n}$) with more draws, the SE for averages and percents SHRINKS (is divided by $\sqrt{n}$) with more draws.

Practice Problems

1) 100 draws are made at random with replacement from the box containing 5 tickets:

$\boxed{0}\boxed{2}\boxed{3}\boxed{4}\boxed{6}$ **<u>The SD of the box is 2</u>**.

a) The Expected Value for the **average** of the **100** draws is ____________.

b) The Standard Error of the **average** of the **100** draws is ____________.

c) The Expected value for the **percent** of $\boxed{6}$'s in **100** draws is ____________.
(HINT: Change to 0-1 box)

d) The Standard Error for the **percent** of $\boxed{6}$'s in **100** draws is ____________.

Chapter 15—Normal Approximation for Probability Histograms

- Probability histograms represent chance by area.
- With a large enough number of draws, the probability histogram for the sum, average and percent of draws from a box will follow the normal curve, even when the contents of the box do not. *This is the Central Limit Theorem.*
- The more the contents of the box differs from the normal curve, the more draws are needed before the probability histogram for the sum of the draws looks like the normal curve.
- To use the normal curve to figure chances, we must first convert the values to Z scores by subtracting the EV and dividing by the SE.

$$\mathbf{Z = \frac{(Value - EV)}{SE}}$$

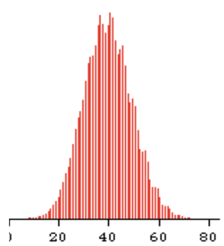
Practice Problems:

1) Look at the 3 boxes and 6 probability histograms below. Each box has 2 probability histograms associated with it. One is the probability histogram for the sum of 2 draws made at random with replacement and the other is the probability histogram for the sum of 20 draws made at random with replacement from the Box.

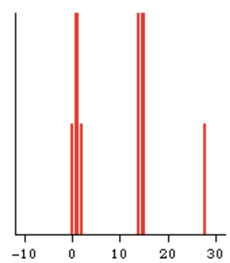
Box A	Box B	Box C
$\boxed{0}\boxed{1}\boxed{2}$	$\boxed{0}\boxed{1}\boxed{5}$	$\boxed{0}\boxed{1}\boxed{14}$

Under each of the 6 histograms, fill in the first blank with either 2 or 20, and the second blank with either A, B or C. The first one is done for you.

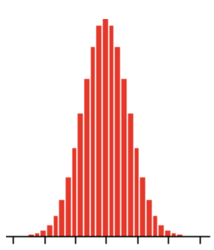
_2_ draws from Box _A_



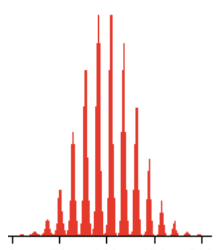
___ draws from Box___



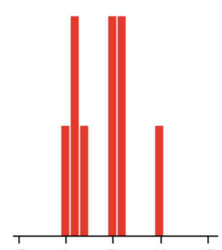
___ draws from Box___



___ draws from Box__



___ draws from Box___



__ draws from Box __

Study Guide for Exam 3

2) A gambler plays roulette **100** times betting \$1 on the numbers 1, 2 and 3 each time. If the ball lands on 1, 2 or 3 the gambler **wins \$11**, if the ball lands on any of the other 35 numbers the gambler **loses \$1**. The roulette wheel has 38 slots numbered 1-36, 0 and 00.

a) What is the appropriate box model? Draw the box.

b) How many draws from the box? __________ (1 pt.)

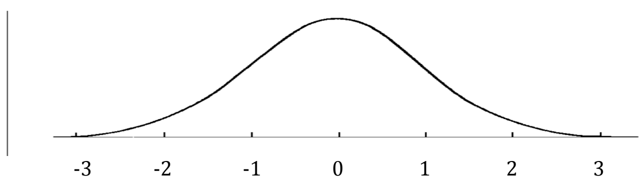
c) What is the average of the box? Write your answer as a fraction. (2 pts.)

d) What is the SD of the box? ***Show work. Circle answer. Round your answer to 2 decimal places.*** (2 pts.)

e) Use the normal approximation and the fact that the **EV is about \$ - 5** and the **SE is about \$32** to figure the chance that the gambler will win more than **\$3** in 100 plays?

i) First calculate the Z score. ***Show work. Circle answer.*** (2 pts)

ii) Now accurately mark z on the normal curve below and ***<u>shade the correct area</u>***. ***Circle your answer.*** (2pts)



3) **100** draws are made at random with replacement from the box containing 5 tickets:
$\boxed{0}\boxed{2}\boxed{3}\boxed{4}\boxed{6}$ **<u>The SD of the box is 2</u>**.

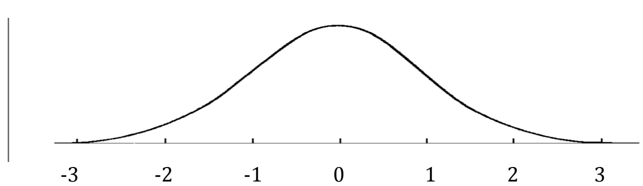
a) The Expected Value of the **average** of the **100** draws is ___________.

b) The Standard Error of the **average** of the **100** draws is ___________.

c) Use the normal approximation and your answers from (a) and (b) above to find the chance that the average of the draws will be more than 3.4.

i) First calculate the Z score. ***Show work. Circle answer.*** (2 pts)

ii) Now accurately mark z on the normal curve below and ***<u>shade the correct area</u>***. ***Circle your answer.*** (2pts)



Chapter 16 – Sample Surveys

- **Random Surveys (subjects are randomly selected) are best for 2 reasons:**
  1. **Eliminate selection bias**
  2. **May be translated into box models so SE's can be calculated.**

- **Self-selected surveys (surveys posted on internet where anyone can respond)**
  have the worst selection bias.

- **Non-response bias and bias in the wording of question is a possible problem for all types of surveys.**

1) Does a simple random sample eliminate the following sources of bias? ***Circle "Yes" or "No".***

a) Selection bias *i)* Yes *ii)* No
b) Non-response bias (not everyone who is selected responds.) *i)* Yes *ii)* No
c) Response bias (response bias deals with how the questions are phrased). *i)* Yes *ii)* No

2) A Fox News poll asked a **randomly** selected sample of 1,015 adults nationwide the following question:
"Do you personally believe in the existence of Hell?"
I asked this class the same question on Bonus Survey 3 last semester. Here are the results.

	Yes	No	Sample Size
Fox random poll	74%	26%	1,015
Bonus Survey 3	54%	46%	546

a) Which survey gives a better estimate of the percentage of **all US adults** who would say "YES" they believe in hell?
Circle one:
i) The Fox survey because the sample was randomly selected from the entire US adult population.
ii) The Fox survey because the sample size is larger.
iii) The Bonus Survey because it was anonymous.
iv) The Bonus Survey because people who watch Fox news are not representative of the population as a whole.

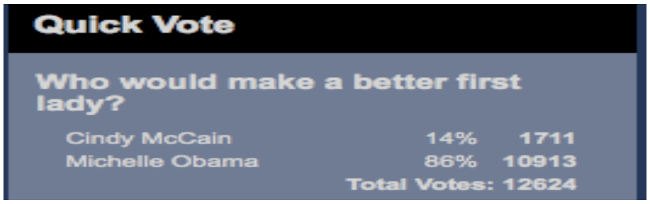
b) The SE of the percentage of people in the Fox sample who answered, "YES" is closest to…
Choose one:

i) $\sqrt{1{,}015} * \sqrt{.74 * .26} * 100\%$ **ii)** $\frac{\sqrt{.74 * .26}}{\sqrt{1{,}015}} * 100\%$

iii) Impossible to compute a SE for this sample because it can't be translated into a box model.

3) Right before the last election www.cnn.com conducted a "quick vote" where they posted the question below on their website and allowed all internet users to cast their votes. Here are the results.

:



a) If we wanted to use this survey to estimate how all US adults would have answered this question at the time what most closely resembles the relevant box model?

Circle one:

i) It has 12,624 tickets; 14% marked "Cindy" and 86% marked "Michelle"

ii) It has millions of tickets marked with "1"s and "0"s. The exact percentages are unknown but are estimated from the sample.

iii) A box model is not appropriate for this poll because the sample was not randomly selected.

b) Does the sample have selection bias? *Circle one:* i) Yes ii) No

c) What is the SE for the percent of all US adults who would answer "Cindy McCain"?

i) $\frac{\sqrt{(.14)(.86)}}{\sqrt{12{,}624}} \, x 100\%$ ii) $\sqrt{12{,}624} \, x \sqrt{(.14)(.86)}$ iii) Impossible to compute a SE for this sample.

Chapter 17- Using Sample Percents to Estimate Population Percents and Building Confidence Intervals around the estimate.

- The sample percent is used to estimate population percent and a margin of error called a confidence interval is attached to the estimate.

- Only the size of the sample, **not the size of the population**, determines the accuracy of the estimate.

- In general, the interval (sample percent $\pm$ **z** $\times$ **SE$_{\%}$**) is a ___**%** confidence interval for the population percentage, where the **blank is the area between z and –z on the Normal Curve.**

  68% Confidence Interval = Sample % +/- **1** × SE$_{\%}$
  95% Confidence Interval = Sample % +/- **2** × SE$_{\%}$
  99% Confidence Interval = Sample % +/- __ × SE$_{\%}$
  80% Confidence Interval = Sample % +/- __ × SE$_{\%}$

- Confidence Intervals can only be applied to the population that the sample was drawn from and NOT to a wider population or to sub-groups within the population.

- Quick formula for choosing how many people to poll for margins of errors for 95% Confidence Intervals: **$N=(100/M_{of}E)^2$**

For a 10% margin of error you need $(100/10)^2 = 100$ people.

For a 5% margin of error you need $(100/5)^2 = 20^2 = 400$ people.

For a 4% margin of error you need $(100/4)^2 = 25^2 = 625$ people.

For a 2% margin of error you need $(100/2)^2 = 50^2 = 2500$ people.

Practice Problems

1. 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".

b) 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:

a) (34.5%-37.5%) **b)** (30%-36%) **c)** (32%-40%) **d)** (35.985%-36.015%)

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

a) be multiplied by 4 **b)** double **c)** be divided by 4 **d)** be divided by 2 **e)** not possible to calculate

2. A CBS News Poll conducted October 19-14, 2011 asked 1,650 ***randomly*** selected US adults nation-wide whether or not they thought money and wealth in this country should be distributed more evenly or whether it is fair the way it is now. 66% of the 1,650 surveyed answered, that it should be distributed more evenly.

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 1,650 tickets, 66% marked "1"and 34% marked"0".
- **ii)** It has millions of tickets, with an average of 0.66, 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, 66% marked "1"and 34% marked"0".

c) Calculate the SE of the sample percent. Show work, circle answer.
(Round to 2 decimal places.)
i) First find the SD of the sample.

ii) Then use it to estimate the SD of the whole population in the $SE_{\%}$ formula:
$SE_{\%}$ = SD of box/ $\sqrt{n}$ x 100% (***Round your answer to the nearest tenth.***)

d) Your best estimate for the % of all American adults who think money and wealth should be more evenly distributed is _________________, give or take about _______.
(Fill in the first blank with the $EV_{\%}$ and the second with the $SE_{\%}$)

e) You can be 95% sure that if you polled all US adults the % who would answer that money and wealth should be more even distributed would be between 66% +/- ______ which means between ____________% and ____________%. Put lower number first.

f) You can be 99% sure that if you polled all US adults the % who would answer that money and wealth should be more even distributed would be between 66% +/- ______ which means between ____________% and ____________%. Put lower number first.

g) Suppose the poll was only taken in Illinois (instead of nationwide), how should CBC adjust the sample size to keep the same SE?

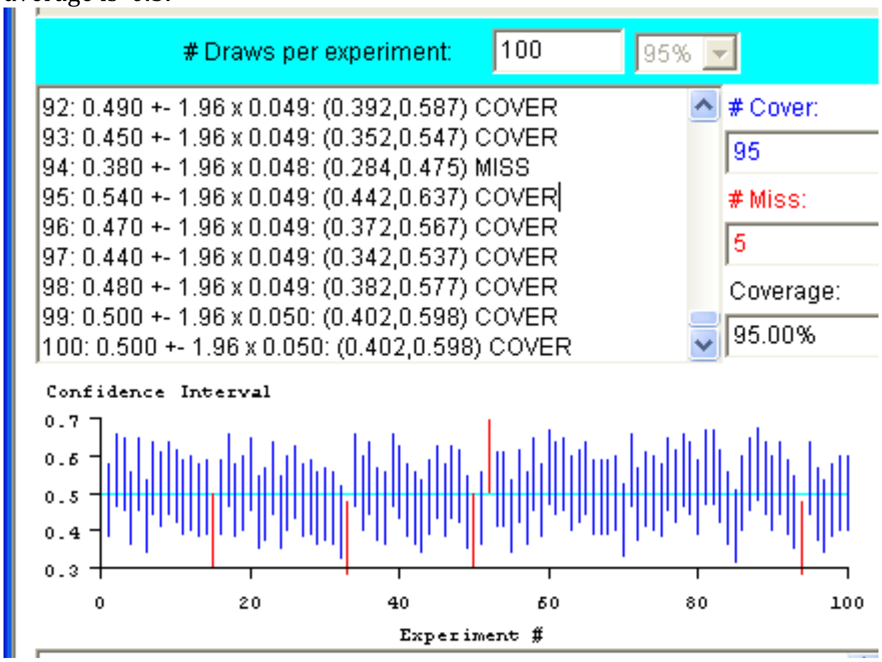
- **i)** Significantly increase sample size
- **ii)** Significantly decrease sample size
- **iii)** Keep sample size about the same.

h) Suppose 1000 pollsters each randomly sampled 1,650 adults nationwide asking whether they thought money and wealth in this country should be distributed more evenly or whether it is fair the way it is now. All 1000 pollsters computed **90%** confidence intervals to estimate the percentage of all US adults who thought money and wealth in this country should be distributed more evenly. Would all 100 intervals correctly include the true population percent?

Choose one:

- **i)** Yes, all 1000 would assuming no errors were made
- **ii)** No, only about 80% of them would
- **iii)** No, only about 90% of them would.

3) 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? ________

b) 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? ________________

c) If each experimenter tossed the coin 400 times, instead of 100 would the width of each confidence interval change, and if so, how? (Hint, how would the SE $_{ave}$ change if n changes from 100 to 400?)

- **i)** It wouldn't change
- **ii)** It would increase by multiplying by 4
- **iii)** It would decrease by dividing by 4
- **iv)** It would increase by multiplying by 2
- **v)** It would decrease by dividing by 2

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

i) 5% of 1000= 50 **ii)** 5 **iii)** $5 * \sqrt{10}$ **iv)** $5/\sqrt{10}$

4) In a pre-election poll in a close race, how many people would you have to poll be **95%** sure that your sample percent is right to within 3%. In other words, how many people do you need to poll to get a margin of error of 3%? (Assume SD = .5)

a) For a 3% margin of error, how many people would you need to poll?

b) How many people would you need to poll get a margin or error within 5%?

Chapter 18- Using Sample Averages to Estimate Population Averages and Building Confidence Intervals around your estimate.

- The sample **average** is used to estimate population average and a margin of error called a confidence interval is attached to the estimate.

- Only the size of the sample, **not the size of the population**, determines the accuracy of the estimate.

- In general, the interval (sample average ± **z** × **SE**$_{\%}$) is a ___**%** confidence interval for the population percentage, where the **blank is the area between z and –z on the Normal Curve.**

68% Confidence Interval = Sample Ave +/- **1** × SE_{ave}
95% Confidence Interval = Sample Ave +/- **2** × SE_{ave}
99% Confidence Interval = Sample Ave +/- __ × SE_{ave}
80% Confidence Interval = Sample Ave +/- __ × SE_{ave}

- Confidence Intervals can only be applied to the population that the sample was drawn from and NOT to a wider population or to sub-groups within the population.

Practice Problems:
1. A polling organization asked a random sample of 1572 adults nationwide this question: "Thinking of 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.)

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

2. A recent poll asked a simple random sample of 400 teenage girls aged 15-19 nationwide how many times in their life they've tanned at a tanning salon. The sample average was 5 times with a SD of 10 times.

a) What most closely resembles the relevant box model? Circle one.

- **i)** It has 400 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 number ranging from 0 to about 1000. The exact average and SD are unknown but are estimated from the sample.
- **iv)** It has 400 tickets. The average of the tickets is 5 and the SD = 10.

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

c) A **95%** confidence interval for the average number of times all US teenage girls aged 15-19 would report tanning in their lifetimes is between __________times and ___________times.
(Put the lower number first.)

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

- **i)** An approximate 95% confidence interval for the average number of times ***all American adults*** tanned at a tanning salon is 4-6 times.
  True False
- **ii)** An approximate 95% confidence interval for the average number of times ***all American teenage girls*** aged 15-19 tanned at a tanning salon is 4-6 times.
  True False
- **iii)** An approximate 95% confidence interval for the average number of times ***all African American teenage girls*** aged 15-19 tanned at a tanning salon is 4-6 times.
  True False
- **iv)** An approximate 95% confidence interval for the average number of times ***all teenage girls world-wide*** aged 15-19 tanned at a tanning salon is 4-6 times.
  True False
- **v)** The sample does not follow the normal curve, so it's impossible to construct any confidence intervals about the population based on our sample.
  True False

e) If the study asked the 400 girls whether or not they had ever tanned at a tanning salon, the relevant box model would contain tickets marked with

- **i)** Only "0"s and "1"s
- **ii)** Numbers ranging from 0 to about 1,00