Statistics 100 Exam 3
Cover Sheet Questions ***(1 pt.)*** Dot Key

1) What's your name? _____ _____
(Last name) (First name)

2) What's your net ID (email)? _____

3) Which Section are you in?
Circle one: **i)** L1 (MWF at noon) **ii)** L2 (TR at 11am) **iii)** ONLINE

****WARNING: When we say "NO WORK, NO CREDIT", we mean it. You'll get a 0.****

Write answers in appropriate blanks. All multiple-choice questions have exactly one answer. If you circle more than one answer you will automatically be marked wrong.

Do NOT use your own scrap paper. Ask a proctor if you need any.

Make sure you have all 6 pages including the Normal table (11 problems).

DO NOT WRITE BELOW THIS LINE _____

The numbers written in each blank below indicate how many points you missed on each page. The numbers printed to the right of each blank indicate how many points each page is worth.

Page 1 _____22

Page 2 _____23

Page 3 _____15

Page 4 _____19

Page 5 _____20

Cover Page_____ 1 *for answering cover page questions 1-3 correctly!*

Total Score _____

There is NO CLASS on Thursday or Friday!
Scores will be posted on Compass by Friday morning and exams returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours on Friday from 4-6 pm or next week.

Page 1- Dot

Statistics 100 Exam 3 April 11, 2018

Question 1 (12 pts.) pertains to these 4 boxes

Box A: [-1] [1] Box B: [-1] [0] Box C: [1] [0] Box D: [1] [1]

a) (4 pts.) Match the boxes above to tossing a fair coin 100 times and counting the following:
(Not all the boxes need to be used and boxes may be used twice.)

i. The number of heads ***Choose one:*** Box A Box B (Box C) Box D

ii. The number of tails ***Choose one:*** Box A Box B (Box C) Box D

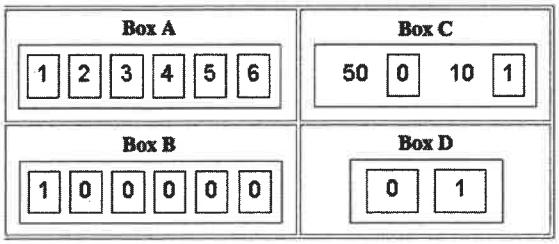
iii. The number of heads minus the number of tails ***Choose one:*** (Box A) Box B Box C Box D

iv. The number of heads plus the number of tails. ***Choose one:*** Box A Box B Box C (Box D)

b) (8 pts.) Fill in the chart below for **each of the boxes.**

Box	Average of box	SD of box	EV_{sum} of 100 draws	SE_{sum} of 100 draws
A	0	1	0	10
B	-0.5	0.5	-50	5
C	0.5	0.5	50	5
D	1	0	100	0

Question 2 (6 pts.) pertains to the 4 box models below:



a) (3 pts.) Rolling a fair die 60 times and counting the number "1's" you get is like drawing 60 times with replacement from Box B.

b) (3 pts.) Taking a random sample of 10 different students from a class that has 50 boys and 10 girls and counting the number of girls you get is like drawing 10 times without replacement from Box C.

For (a) and (b) above, fill in the first blank with a number, the second with either "with" or "without", and the third with the correct Box.

Question 3 (4 pts.) **pertains to tossing a fair coin and the law of averages.**
Do the 2 scenarios have the same probability or not?

a) (2 pts.) The probability of getting 50% ± 10% heads in **10** tosses ____the probability of getting 50% ± 1% heads in **1,000** tosses
Circle one: **i)** is equal to **ii)** is not equal to

b) (2 pts.) The probability of getting 5 ± 1 heads in **10** tosses ____ the probability of getting heads 500 ± 1 in **1,000** tosses
Circle one: **i)** is equal to **ii)** is not equal to

Question 4 (23 pts.) **25** draws are made at random with replacement from the box containing 6 tickets: 1 5 5 7 7 11 **(SD of box = 3)**

a) (2 pts.) The ***smallest*** the sum of the 25 draws could possibly be is 25 and the ***largest*** is 275.

b) (2 pts.)What is the **EV** of the **sum** of the **25** draws? (Show work, circle answer.) **No work, no credit.** ✱

$EV_{sum} = n \times \text{avg of box} = 25 \times 6 = (150)$

c) (2 pts.)What is the **SE** of the **sum** of the **25** draws? (Show work, circle answer.) **No work, no credit.** ✱

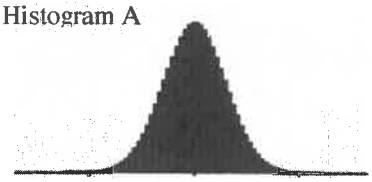
$SE_{sum} = \sqrt{n} \times \text{SD of box} = \sqrt{25} \times 3 = (15)$

d) Look at the 3 probability histograms below. One shows the **contents of the box**, one shows the **sum of 2 draws** with replacement from the box and one shows the **sum of 25 draws** with replacement for the box. Which is which?

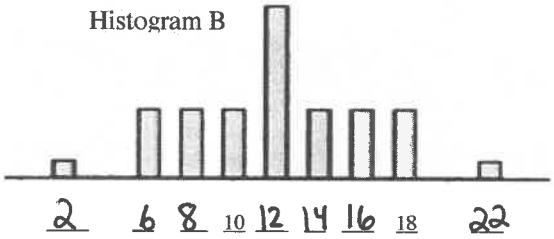
i) (3 pts.)Histogram B is for the sum of 2 draws, Histogram A is for the sum of 25 draws and Histogram C is for the contents. (***Fill in the 3 blanks with A, B, or C***)

ii) (4 pts.) The numbers on the X axes in the histograms are missing. What numbers belong under each of the 7 remaining bars in Histogram B and under each of the 4 bars in Histogram C? (Two numbers are already done for you in Histogram B.)

Histogram A

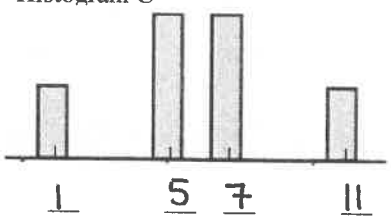


Histogram B



(2 pts.) Write the correct number in each blank

Histogram C



(2 pts.) Write the correct number in each blank.

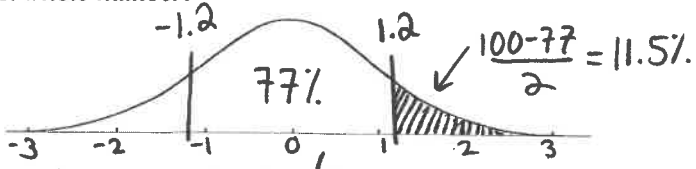
e) (3 pts.) The **EV** of the **sum** of **100** draws from the box = **600** and the **SE** of the **sum** of **100** draws = **30**. Use the normal approximation to find the **chance** that the sum of **100 draws** will be **above 636**?

i) (2 pts.) First calculate the Z score. ***Show work. Circle answer.*** **No work, no credit.** ✱

$Z = \frac{\text{val} - EV}{SE} = \frac{636 - 600}{30} = (1.2)$

ii) (1 pt.) Now use the normal table to find the chance that the sum of the 100 draws will be ***above 636.*** ***Round the middle area given in the table to the nearest whole number.***

Chance = 11.5 % (1 pt.)
Round answer to 1 decimal place.



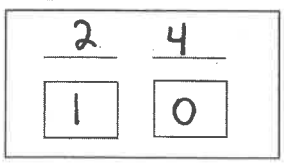
f) (1 pt.)What is the **EV** of the **average** of the **100** draws from the same box above? 6

g) (2 pts.) What is the **SE** of the **average** of the **100** draws? 0.3 **Show work. No work, no credit.** ✱

$SE_{avg} = \frac{\text{SD of box}}{\sqrt{n}} \times \frac{3}{\sqrt{100}} = 0.3$

h) (4 pts.) Now suppose you draw at random with replacement from the same box above, but this time you're only interested in the percent of 7's. What is the **EV and SE** of the percent of 7's in **100** draws?

i) (2 pts.) Draw a new box. **Label the two tickets with the correct numbers, and write how many of each above them.**



ii) (1 pt.) $EV_{\%}$ of 7's in **100** draws = 33.33 % ***(Round to 2 decimal places.)***

iii) (1 pt) $SE_{\%}$ of 7's in **100** draws is … ***Show work below.***
Circle one: **a)** 0.37% **b)** 0.47% **c)** 0.5% **d)** 3.7% **(e)** 4.7% **f)** 5%

$SE_{\%} = \frac{\text{SD of box}}{\sqrt{n}} \times 100 = \frac{|1-0|\sqrt{2/6 \times 4/6}}{\sqrt{100}} \times 100 = 4.7\%$

Question 5 (9 pts.) **pertains to the following situation:**
In roulette, there are 18 red numbers, 18 black numbers and 2 green numbers. Consider betting $1 on "Black." If black comes up, you win $1, but if black does not come up, you lose $1.
The average of the corresponding box is $-.05, and the SD of the box is $1. Imagine playing this bet 100 times.

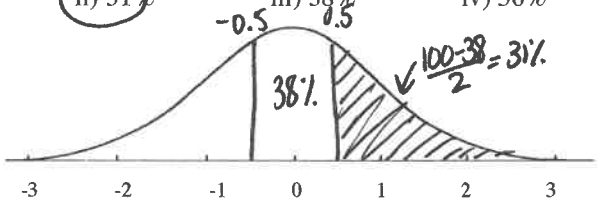
a) (2 pts.) The amount of money you get from playing this bet 100 times is like drawing from what box?
i) The box has two tickets: 1 marked "1" and 1 marked "-1".
ii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
iii) The box has 38 tickets: eighteen 1's, eighteen -1's, and two 0's
iv) The box has 100 tickets: half are -1's, half are 1's.
(v) The box has 38 tickets: eighteen are 1's and twenty are -1's

b) (2 pts.) The expected value of the sum of the 100 draws is closest to
i) $5 (ii) $ -5 iii) $0 iv) $0.50 v) $ - 0.50

c) (2 pts.) The standard error of the sum of the 100 draws is closest to
i) $1 ii)$ 5 (iii) $10 iv) $0.10 v) $100

d) (2 pts.) Use the normal curve to estimate the chance that you come out ahead. In other words, estimate the chance that the sum of the 100 draws is more than $0. Circle the closest answer and **show work for credit.**

i) 20% (ii) 31% iii) 38% iv) 50% v) 69%



$Z = \frac{val-EV}{SE} = \frac{0-(-5)}{10} = 0.5$

e) (1 pt.) If you play the game 400 times, will the chance that you come out ahead be larger than, smaller than or the same as when you play 100 times? (i) smaller than ii) larger than iii) the same as

Question 6 (6 pts.) **pertains to the following situation:**
During the same week in March, 3 polls asked this question: "Do you think arming teachers will prevent more mass shootings?" The Quinnipiac poll asked the question of a ***randomly*** selected sample of 1,112 adults nationwide. The MSNBC Live Live poll posted the question on its website and allowed anyone who visited their website to cast a vote, and you were asked the question on our Bonus Survey 3. Here are the results:

	Yes	No	Sample Size
Quinnipiac random poll	40%	60%	1,112
MSNBC Live Vote	53%	47%	6,152
Bonus Survey 3	14%	86%	991

a) (1 pt.) Which poll best reflects how **all US adults** would answer this question? ***Choose one:***
(i) The Quinnipiac Poll because the sample was ***randomly*** selected from the entire US adult population.
ii) The MSNBC Live Vote poll because it has the largest sample size.
iii) The Bonus Survey 3 poll since it has the lowest non-response bias.

b) (3 pts.) For each poll listed below, is it possible to calculate a 95% Confidence Interval for the % of **ALL US adults** who would answer "Yes" to this question? ***Circle either "Yes" or "NO" for each poll.***

i)	Quinnipiac random Poll	(Yes)	or	No
ii)	MSNBC Live Poll	Yes	or	(No)
iii)	Bonus Survey 3	Yes	or	(No)

c) (2 pts.) What is SE of the sample percent for the Quinnipiac poll? ***Choose one:***
i) It's not possible to calculate a SE because we don't know the SD of the sample.
ii) It's not possible to calculate a SE because we don't know the exact size of the population.
(iii) The SE of the sample percent is approximately 1.5%
iv) The SE of the sample percent is approximately 3%
v) The SE of the sample percent is approximately 5%

$SE_{\%} = \frac{|1-0|\sqrt{0.40 \times 0.60}}{\sqrt{1112}} \times 100$

Question 7 (9 pts.) **pertains to the following situation:**
To estimate the average weight of all 15,000 UI male undergraduates a simple random sample of size 100 is picked. Suppose the weights of the 100 men in the sample follow the **normal curve** with an **average =160 lbs. with a SD = 20 lbs.**

a) (2 pts.) What most closely resembles the relevant box model? ***Circle one.***

i) It has 100 tickets marked with "0"s and "1"s.
ii) It has about 15,000 tickets marked with "0"s and "1"s..
(iii) It has about 15,000 tickets. On each ticket is written a weight. The exact average and SD are unknown.
iv) It has millions of tickets. On each ticket is written a weight. The average of the tickets is 160 lbs. and the SD = 20 lbs.

b) (2 pts.) We can estimate the **average weight** of all male UI undergraduates to be 160 lbs. give or take ____ lbs. or so. (Circle the SE for the sample average.) **Show work for credit.**

i) 0 .2 ii) 20 (iii) 2 $SE_{avg} = \frac{SD\ of\ box}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$

c) (5 pts.) Circle whether each of the statements below are true or false:

i)	160 ± 4 lbs. is a 95% confidence interval for the **average** weight of all US males.	True	(False)
ii)	About 95% of the students in the sample weigh between 160 ± 4 lbs.	True	(False)
iii)	About 95% of the students in the sample weigh between 160 ± 40 lbs.	(True)	False
iv)	About 95% of all UI male undergraduates weigh 160 ± 4 lbs.	True	(False)
v)	160 ± 4 lbs. is a 95% confidence interval for the **average** weight of all male UI undergraduates.	(True)	False

Question 8 (10 pts.) **pertains to the following situation:**
A recent Fox News Poll asked a random sample of 900 adults nationwide the following question: "Have you ever cheated on a person you were in a relationship with?" 21% of the people in the sample answered "YES".

a) (2 pts.)What most closely resembles the relevant box model? Circle one.

i) It has 900 tickets, 21% marked "1"and 79% marked"0".
ii) It has millions of tickets, with an average of 0.21, 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, exactly 21% marked "1"and exactly 79% marked"0".

b) (1 pt) The draws are made _____replacement. i) With (ii) Without

c) (2 pts) What is the expected value for the percent of all US adults who would say they cheated in a relationship? 21 %

d) (1 pt) What is the expected value for the percent of all US adults who ***actually*** cheated in a relationship?

i) 21% ii) 79% (iii) not enough information to determine

e) (2 pts.) Calculate the SE for the percentage of people in the sample who answered "YES". Show work, circle answer. (Round to 2 decimal places.) **No work, no credit.**

$SE_{\%} = \frac{SD}{\sqrt{n}} \times 100 = \frac{|1-0|\sqrt{.21 \times .79}}{\sqrt{900}} \times 100 = (1.36\%)$

f) (2 pts.) Suppose the pollsters wanted divide the SE from part (e) by 2, they would need to multiply the sample size by 4.
Fill in the first blank with either "multiply" or "divide" and the second blank with a number

Question 9 (6 pts.)

A Fox News Poll asked a random sample of 500 men and 500 women nationwide the following question: "Trying to be as honest as you can, are you more attracted to people by their bodies or their brains?" 24% of the women and 44% of the men in the samples answered "Bodies".

a) (2 pts.) An approximate **68%** confidence interval for the percentage of all American men who would say they are more attracted to bodies than brains is:

(i) $44\% \pm \frac{\sqrt{0.44*0.56}}{\sqrt{500}} * 100\%$ **ii)** $44\% \pm \frac{\sqrt{0.44*0.56}}{\sqrt{1000}} * 100\%$ **iii)** $68\% \pm \frac{\sqrt{0.44*0.56}}{\sqrt{500}} * 100\%$ **iv)** $44\% \pm \frac{\sqrt{0.24*0.44}}{\sqrt{1000}} * 100\%$

b) (2 pts.) The sample sizes for the men and women are exactly the same. Does that mean the $SE_{\%}$ of the 2 samples are also the same?

i) Yes ii) No, the male $SE_{\%}$ is smaller (iii) No, the female $SE_{\%}$ is smaller iv) Not enough info

c) (2 pts.) If you combine the 2 samples into one sample of 1000 you get 34% answering "Bodies". How does the SE of the combined sample compare to the SE's of the separate male and female samples?

i) It's bigger (ii) It's smaller iii) It's the same iv) It's smaller than one and bigger than the other v) Not enough info

Question 10 (8 pts.)

A Harris Poll asked a random sample of **900** adults nationwide the following question: "Are you afraid of the dark?" **28%** of the people in the sample answered "YES".

a) (2 pts.) What is the SD of the sample? ***Show work.*** **No work no credit.**

$SD = |1-0|\sqrt{.28 \times .72} = 0.449$

SD= 0.449 (2 pts.)

Round answer to 3 decimal places

b) (2 pts.) An approximate **80%** confidence interval for the percentage of all American adults who would answer that they are afraid of the dark is….. ***Choose one:***

i) $28\% \pm \mathbf{2} * (SE_{\%})$ (ii) $28\% \pm \mathbf{1.3} * (SE_{\%})$ iii) $28\% \pm \mathbf{0.8} * (SE_{\%})$ iv) $80\% \pm \mathbf{1.3} * (SE_{\%})$

c) (2 pts.) If the researcher increased the sample size to 3600 then the width of the confidence interval above would…***Choose one:***

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

d) (2 pts.) Suppose 200 pollsters each randomly sampled 900 adults nationwide asking whether they were afraid of the dark. All 200 pollsters computed **80%** confidence intervals to estimate the percentage of all US adults who would answer that they were afraid of the dark. About ***how many*** of the 200 confidence intervals would miss the true population percentage? 40

(No credit will be given if you answer the percent instead of the number.)

$200 \times 0.20 = 40$

Question 11 (6 pts.)

Suppose a survey organization is planning to take a random poll in 2 cities to estimate the percentage of people who would support arming teachers in their city. City A has a population of 1 million and City B has a population of 16 million.

a) (2 pts.) Other things being equal, to achieve the same level of accuracy in the both polls, the number of people you'd have to poll in City A is about _____ the number of people you'd have to poll in City B. ***Choose one:***

i) 16 times larger than **ii)** 4 times larger than **(iii)** the same as **iv)** 4 times smaller than **v)** 16 times smaller than

b) (2 pts.) How many people would you have to poll in City A to get a 95% Confidence Interval with a Margin of Error of **4%?** (Assume the SD of the population is close to 0.5)

i) 400 (ii) 625 iii) 1111 iv) 2500 v) 10,000

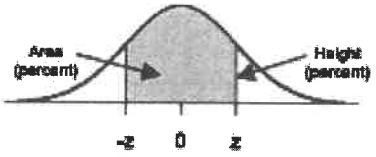
$n = \left(\frac{200 \times 0.5}{4}\right)^2 = 625$

c) (2 pts.) How many people would you have to poll in City B to get a 95% Confidence Interval with a Margin of Error of **10%?** (Assume the SD of the population is close to 0.5)

(i) 100 ii) 400 iii) 625 iv) 2500 v) 10,000

$n = \left(\frac{200 \times 0.5}{10}\right)^2 = 100$

STANDARD NORMAL TABLE



z	Area		z	Area		z	Area
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991