Statistics 100 Exam 3
Cover Sheet Questions *(1 pt.)* Comma 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 (TR at 12:30 pm) **ii)** L2 (MWF at noon) **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 (12 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 _____20

Page 2 _____32

Page 3 _____21

Page 4 _____16

Page 5 _____10

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

Total Score _____

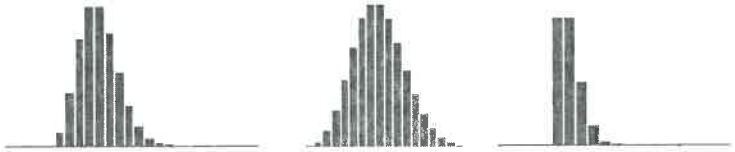
There is NO CLASS on Thursday or Friday!
Scores will be posted on Compass by Friday morning and exams returned in class after break. Online students may pick up their exam in 23 Illini Hall during office hours on Friday from 2:30-4 pm or after break.

Statistics 100 Exam 3 Page 1- Comma November 14, 2018

Question 1 (6 pts.)

The 3 histograms below (in scrambled order) are the probability histograms for the sum of **100, 400 and 900** random draws with replacement from a box that has 99 tickets marked "0" and only 1 marked "1": 99 [0] 's [1]

Histogram A Histogram B Histogram C



Fill in the blanks below to match the histograms with the correct number of draws. (Hint: You do NOT need to see the histograms' horizontal axis to answer these questions. Just compare the ***shape*** of the histograms to the number of draws.)

i) *(2 pts.)* Histogram C is the probability histogram for the sum of **100** draws from the box.

ii) *(2 pts.)* Histogram A is the probability histogram for the sum of **400** draws from the box.

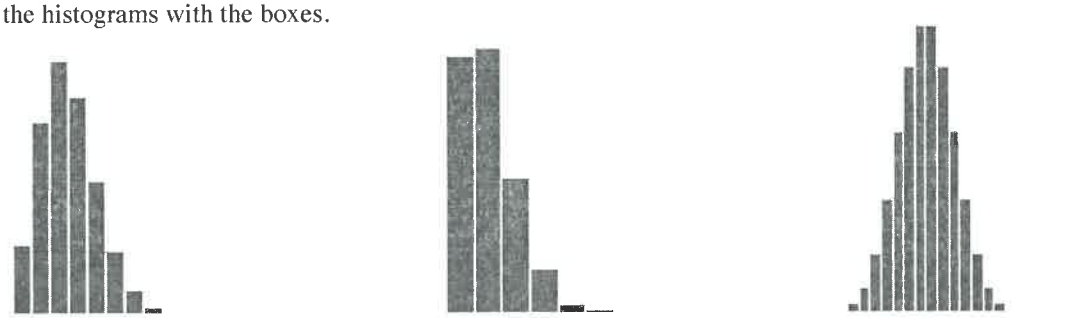
iii) *(2 pts.)* Histogram B is the probability histogram for the sum of **900** draws from the box.

Question 2 (6 pts.)

25 draws are made with replacement from each of the following boxes:

Box A [0] [1] Box B 9 [0] 's [1] Box C 24 [0] 's [1]

The probability histograms for the sums of 25 draws taken from each box are shown below, in scrambled order. Match the histograms with the boxes.



(2 pts.) **Box** B *(2 pts.)* **Box** C *(2 pts.)* **Box** A

Fill in the blanks with A, B or C.

(HINT: You do NOT need to see the histogram's horizontal axis to answer these questions. Just compare the *shapes* of the histograms to the contents of each box.)

Question 3 (8 pts.) pertains to these 6 boxes:

Box A	Box B	Box C	Box D	Box E	Box F
1 2 3 4 5 6	2 3 4 5 6 7 8 9 10 11 12	1 0	0 0 0 0 1	0 0 0 0 0 1	1 -1

a) (4 pts.) Match the boxes above to the following scenarios. (***Not all the boxes will be used and boxes may be used twice.)***

i. A fair die is rolled 100 times and the number of 2's is counted. Box E

ii. A fair pair of dice is rolled once and the total number of spots is counted. Box A

iii. A fair coin is tossed 50 times and number of heads minus the number of tails counted. Box F

iv. A multiple-choice test has 100 questions. Each question has 5 answers (only 1 of which is right). Suppose you guess at random on each question and the number of correct answers is counted. Box D

b) (4 pts.) Look at Boxes C, D, and E.

i) *(2 pts.)* Which has the largest SD? Circle one: (C) D E

ii) *(2 pts.)* Which has the smallest SD? Circle one: C D (E)

Page 2 - Comma

Statistics 100 Exam 3 November 14, 2018

Question 4 pertains tossing a fair coin: (6 pts.)

a) (4 pts.) A coin is tossed 100 times and EV_{sum} = 50 heads and SE_{sum} = 5 heads and the $EV_{\%}$=50% and $SE_{\%}$=5%.
Now suppose you toss the coin 400 times then the EV_{sum} = 200 heads and SE_{sum} = 10 heads and $EV_{\%}$= 50 % and $SE_{\%}$= 2.5 %.
Fill in the 4 blanks with numbers.

$z = \frac{190-200}{10} = -1$

b) (2 pts.) The chance of getting between 190 and 210 heads in 400 tosses of a fair coin is closest to

i) 68% ii) 95% iii) 80% iv) 5%

$z = \frac{210-200}{10} = 1$

Question 5 (26 pts.) **100** draws are made at random with replacement from the box containing 4 tickets: 1 3 3 9. avg = 4
(SD of box = 3)

a) (2 pts.)What is the **EV** of the **sum** of the **100** draws? (Show work, circle answer.) **No work, no credit.**

EVsum = n × avg of box = 100 × 4 = (400)

b) (2 pts.)What is the **SE** of the **sum** of the **100** draws? (Show work, circle answer.) **No work, no credit.**

SEsum = $\sqrt{n}$ × SD of box = $\sqrt{100}$ × 3 = (30)

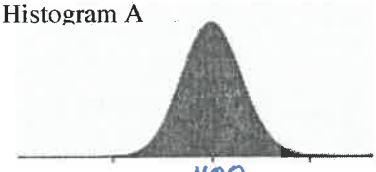
c) (10 pts.) Look at the 3 probability histograms below. One shows the **contents of the box**, one shows the **sum of 2 draws** with replacement from the box and one shows the **sum of 100 draws** with replacement for the box. Which is which?

i) (3 pts.)Histogram C is for the sum of 2 draws, Histogram A is for the sum of 100 draws and Histogram B is for the contents of the Box. ***(Fill in the 3 blanks with A, B, or C)***

more bars ⇒ more draws

ii) (7 pts.) The numbers on the X axes in the histograms are missing. What number belongs in the middle of Histogram A and what numbers belong under each of the 3 bars in Histogram B and under each of the 6 bars in Histogram C?

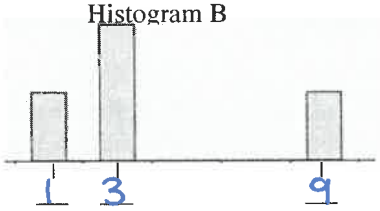
Histogram A



Write the correct number in the blank. (1 pt)

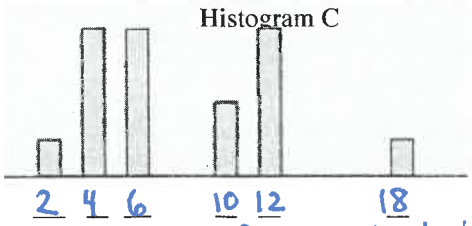
EV in middle of normal curve

Histogram B



Write the correct numbers in each blank. (3 pts.)

Histogram C



outcomes for drawing twice from box

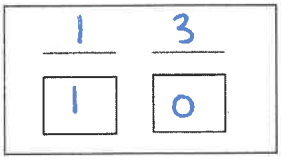
Write the correct number in each of the 6 blanks. (3 pts.)

d) (6 pts.) The **EV** of the **average** of **100** draws from the box = 4 and the **SE** of the **average** of **100** draws = 0.3.
There's a 95% chance that the average of the100 draws will be between 3.4 and 4.6.
Fill in the 4 blanks with the correct numbers. 95% CI = 4 ± 2(0.3)

$SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{3}{\sqrt{100}}$

e) (6 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 9's. What is the **EV and SE** of the percent of 9'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.**



4 tickets

ii) (2 pt.) $EV_{\%}$ of 9's in **100** draws = 25 %.

iii) (2 pt) $SE_{\%}$ of 9's in **100** draws is … ***Show work below.***
Circle one: **a)** 0.37% **b)** 0.43% **c)** 0.5% **d)** 3.7% **(e)** 4.3% **f)** 5%

$SE_{\%} = \frac{SD \text{ of box}}{\sqrt{n}} \times 100\%$

$= \frac{|1-0|\sqrt{\frac{1}{4} \times \frac{3}{4}}}{\sqrt{100}} \times 100 = 4.33$

Statistics 100 Exam 3 November 14, 2018

Question 6 pertains to the following situation: (15 pts.)

In roulette, there are 18 red numbers, 18 black numbers and 2 green numbers. Consider betting $1 on "Red." If red comes up, you win $1, but if red does not come up, you lose $1. The average in the corresponding box is -.0526, and the SD in 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) It has two tickets: 1 marked "1" and 1 marked "-1".
ii) It has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
(iii) It has 38 tickets: eighteen are 1's and twenty are -1's
iv) It has 400 tickets: half are -1's, half are 1's.
v) It has 38 tickets: eighteen 1's, eighteen -1's, and two 0's

$\text{avg} = \frac{18(1)+20(-1)}{38} = -\frac{2}{38}$

$SD = |1-(-1)|\sqrt{\frac{18}{38}\times\frac{20}{38}} = 1$

b) *(1 pt.)* With or without replacement? (i) with ii) without

c) *(2 pts.)* The expected value of the sum of the draws is closest to $EV_{sum} = n \times avg = 100 \times -2/38 = -5.26$

i) $5.26 ii) $-0.526 iii) 0 iv) $0.526 (v) $ -5.26

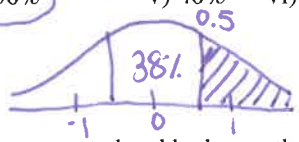
d) *(2 pts.)* The standard error of the sum of the draws is closest to $SE_{sum} = \sqrt{n} \times SD = \sqrt{100} \times 1 = 10$

i) $1 ii)$ 5.26 (iii) $10 iv) $0.10 v) $100

e) *(2 pts.)* Use the normal curve to estimate the chance that you come out ahead (i.e. that the sum of the draws is more than 0.) The answer is closest to

i) 10% ii) 15% iii) 20% (iv) 30% v) 40% vi) 50%

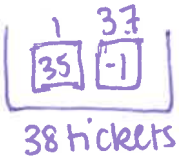
$Z = \frac{val - EV}{SE} = \frac{0-(-5.26)}{10} = 0.5$



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

f) *(2 pts.)* 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

g) Let's consider a different roulette bet—betting $1 on the number "7". If the ball lands on "7" you win $35, if it lands on any of the other numbers you lose $1. The box model for this bet would be one ticket marked $35 and 37 tickets marked $-1.



i) *(2 pts.)* What is the average of the box? Show work. No work, no credit. Average = $-2/38$ (Write answer as fraction.)

$\text{avg} = \frac{1(35)+37(-1)}{38} = -2/38$

ii) *(2 pts.)* What is the SD of the box? Show work. No work no credit. SD = 5.76 (Round to 2 decimal places.)

$SD = |35-(-1)|\sqrt{1/38 \times 37/38} = 5.76$

Question 7 (4 pts.)

A Fox News Poll asked a random sample of 900 adults nationwide the following question: "Do you think there will ever be peace in the Middle East?" Only 17% of the sample answered "YES".

a) *(2 pts.)* The EV of the percent of people in the nation would answer "Yes to this question is 17 %.

b) *(2 pts.)* What is the $SE_{\%}$?

(i) $\frac{\sqrt{(.17)(.83)}}{\sqrt{900}} \times 100\%$ ii) $\sqrt{900} \times \sqrt{(.17)(.83)}$ iii) Impossible to compute a SE for this sample.

Question 8 *(2 pts.)*

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

a) quite a bit lower than b) about the same as (c) quite a bit higher than

sample in NY > sample in Montana
bigger sample ⇒ more accurate

Question 9 *(2 pts.)*
A poll is taken in a city of population 200,000. A simple random sample of size 1000 is chosen and polled. Another poll is to be taken in the same way in a second city of population 400,000. In order to obtain the same accuracy as in the first city, the sample size in the second city should be:

Keep sample size the same

a) 1000 b) 2000 c) 4000 d) 8000

Question 10 pertains to the following situation: (10 pts.)
A Gallup poll conducted March 5-7, 2001, asked a simple random sample of 1060 adults nationwide whether they thought the government should "set legal limits on the amount of energy which average consumers can use".
35% of the sample favored setting legal limits.

a) *(2 pts.)* What most closely resembles the relevant box model (the box from which the tickets are drawn)? Circle one.
i) It has 1060 tickets, 35% are marked "1" and 65% are marked "0"
ii) It has 1060 tickets with an average of 0.
iii) It has millions of tickets marked "0" and "1", but the exact percentage of each is unknown.

b) *(2 pts.)* The draws are made _____ replacement.
i) With ii) Without

c) *(2 pts.)* The SE of the sample percent is about 1.5% An approximate 95% confidence interval for the percentage of all Americans who favor setting legal limits on consumer energy use is $95\%\ CI = 35\% \pm 2(1.5\%)$

i) (33.5%-36.5%) ii) (32%-38%) iii) (30%-40%) iv) (32.7%-35.3%)

d) *(2 pts.)* Suppose 80 pollsters each randomly sampled 1060 adults nationwide asking them whether the government should set legal limits on energy use. All 80 pollsters computed **90%** confidence intervals to estimate the percentage of all US adults who would favor setting legal limits on energy use. About ***how many*** of the 80 confidence intervals would miss the true population percentage? 8 -1 for answer = 72

10% of 80 would miss
$.10 \times 80 = 8$

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

e) *(2 pts.)* If the sample size of the poll was increased by a factor of 4 (to n = 4240) then the width of the 95% confidence interval would _____

$n \uparrow$, SE% $\downarrow$ by $\sqrt{n}$

i) increase by a factor of 2 ii) increase by a factor of 4 iii) decrease by a factor of 2 iv) decrease by a factor of 4

Question 11 (4 pts)
a) *(2 pts.)* A survey organization wants to take a simple random sample in a city with a population of 400,000 in order to estimate the percentage of adults in who will vote in the next presidential election. (Assume the percentage is similar to what it's been in the past, about 50%). About how many people would they have to poll to get a 95% Confidence Interval with a Margin of Error of **4%?**

i) 400 ii) 625 iii) 1111 iv) 2500 v) 10,000 $n = \left(\frac{200 \times 0.5}{4}\right)^2 = 625$

b) *(2 pts.)* If the city had a population of only 100,000 how should the survey organization adjust the sample size to keep the same margin of error?

i) Increase it by a factor of 2
ii) Decrease it by a factor of 2
iii) Keep it the same

pop. size doesn't matter

Question 12 pertains to the following survey: (10 pts.)

As part of a survey on alcohol use among UI students, a survey is given to a simple random sample of 400 UI students. In the survey, the students report an average of 7.3 drinks per week with a SD of 9.8 drinks. Assume there are 27,000 undergraduates at the UI.

a) *(2 pts.)* What most closely resembles the relevant box model (the box from which the tickets are drawn)? Circle one.

i) It has 27,000 tickets, 7.3% are marked "1" and 92.7% are marked "0"
ii) It has 27,000 tickets with an average of 0.
iii) It has 27,000 tickets marked with numbers ranging from 0 to about 50, but the exact average and SD is unknown.
iv) It has 27,000 tickets marked with numbers ranging from 0 to about 50; the average of the tickets is 7.3 and the SD is 9.8.

b) *(2 pts.)* The draws are made _____replacement. i) With ii) Without

c) *(2 pts.)* We can estimate the average number of drinks among all UI students to be 7.3 drinks, and the SE for the sample average is closest to …

i) 9.8 ii) 0.5 iii) 1 iv) 0.1

$SE_{avg} = \frac{\text{SD of box}}{\sqrt{n}} = \frac{9.8}{\sqrt{400}} = 0.49$

d) *(4 pts.)* Determine which of the statements below are true and which are false:

i) 7.3 ±1 is an apporximate 95% confidence interval for the average number of drinks all US students consume per week.
True False

ii) About 95% of the students in the survey reported drinking between 7.3 ±1 drinks per week.
True False

iii) About 95% of U. of I. students have between 7.3 ±1 drinks per week.
True False

iv) 7.3 ±1 is an approximate 95% confidence interval for the average number of drinks all UI students consume per week.
True False