Stat 100 Exam 3
EXAM 3: Statistics 100

Key - Form A

Spring 2023

READ THE DIRECTIONS BELOW TWICE!

Cover Sheet Questions

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

2) What's your **net ID** (email)? _____@illinois.edu

3) Which **section** are you in? ***Circle one:***

i) L2 (In Person Section) **ii)** O1 (Online Section)

This test is ALL multiple choice. **Circle all answers on this exam and fill in the corresponding bubble on your orange scantron**. All questions have exactly one answer. If you circle/bubble in more than one answer, you will automatically be marked wrong. Make sure to circle the answers on this test and fill out your scantron. **If you don't do both, you will get a 0.**

SCANTRON Directions

- Print and bubble in your LAST NAME with **no spaces** starting in the left most column. Print your FIRST INITIAL in the right-most column.
- Print and bubble in your UIN number in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box.
- *Write Stat 100 on the COURSE line.*
- *Write your instructor's name (Karle Flanagan) on the INSTRUCTOR line.*
- *Write your section (L2 or O1) on the SECTION line.*
- *Sign your name, and right underneath the student signature line PRINT your name.*

READ THIS: Failure to fill out your scantron correctly will result in a loss of 2 points on your exam!

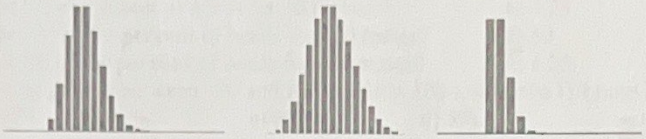
WARNING- The exams look alike but you are sitting next to people who actually have a different version than you. Copying from anyone is equivalent to giving a signed confession.
All cheating including being caught with a non-permissible calculator or formula sheet will result in a 0 and an academic integrity violation on your university record.

There is NO CLASS on Friday this week!

Scores will be posted on Canvas by Monday at noon. Students may pick up their exam in 171 Computing Applications Building during office hours next week.

Questions 1-3 pertain to the following histograms: 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".

Histogram A **Histogram B** **Histogram C**



more bars → more draws

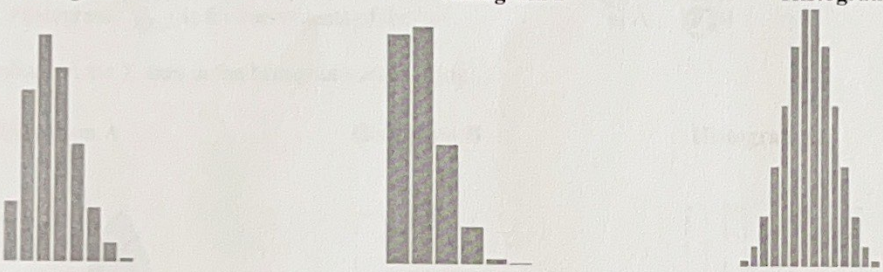
1. Histogram C is the probability histogram for the sum of **100** draws from the box. **a)** A **b)** B **(c)** C
2. Histogram A is the probability histogram for the sum of **400** draws from the box. **(a)** A **b)** B **c)** C
3. Histogram B is the probability histogram for the sum of **900** draws from the box. **a)** A **(b)** B **c)** C

Questions 4-6 pertain to this situation: 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.

Histogram 1 **Histogram 2** **Histogram 3**



4. Histogram 1 corresponds to which box? **a)** A **(b)** B **c)** C
5. Histogram 2 corresponds to which box? **a)** A **b)** B **(c)** C
6. Histogram 3 corresponds to which box? **(a)** A **b)** B **c)** C

Questions 7-12 pertain to these boxes below:

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

Match the boxes above to the following scenarios.

7. A fair die is rolled 100 times and the number of 2's is counted.
   a) Box A **b)** Box B **c)** Box C **(d)** Box D **e)** Box E
8. A fair die is rolled once and the total number of spots is counted.
   (a) Box A **b)** Box B **c)** Box C **d)** Box D **e)** Box E
9. A fair coin is tossed 50 times and number of heads minus the number of tails counted.
   a) Box A **b)** Box B **c)** Box C **d)** Box D **(e)** Box E
10. 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.
    a) Box A **b)** Box B **(c)** Box C **d)** Box D **e)** Box E

Look at Boxes B, C, D, and E.

11. Which has the largest SD? **a)** Box B **b)** Box C **c)** Box D **(d)** Box E
12. Which has the smallest SD? **a)** Box B **b)** Box C **(c)** Box D **d)** Box E

Questions 13-17 pertains tossing a fair coin:

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.

13. What is the EV of the **sum** of heads for 400 tosses? a) 50 b) 100 (c) 200 d) 400
14. What is the SE of the **sum** of heads for 400 tosses? a) 1.25 b) 2.5 c) 5 (d) 10
15. What is the EV of the **percent** of heads for 400 tosses? (a) 50 b) 100 c) 200 d) 400
16. What is the SE of the **percent** of heads for 400 tosses? a) 1.25 (b) 2.5 c) 5 d) 10
17. The chance of getting between 190 and 210 heads in 400 tosses of a fair coin is closest to?
(a) 68% b) 95% c) 80% d) 5%

$200 \pm 1(10)$

Questions 18-28 pertain to this scenario: 100 draws are made at random with replacement from the box containing 4 tickets: 1 3 3 9. The SD of the box is 3.

18. What is the EV of the sum of the 100 draws? a) 4 b) 100 c) 300 (d) 400

$EVsum = n \times avg = 100 \times 4 = 400$

19. What is the SE of the sum of the 100 draws? a) 0.3 b) 3 (c) 30 d) 300

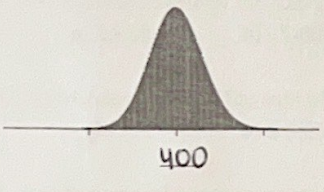
$SEsum = \sqrt{n} \times SD = \sqrt{100} \times 3 = 30$

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 from the box. 1 3 3 9. Which is which?

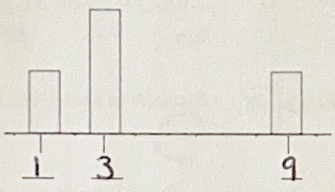
20. Histogram C is for the sum of 2 draws. a) A b) B (c) C
21. Histogram A is for the sum of 100 draws. (a) A b) B c) C
22. Histogram B is for the contents of the box. a) A (b) B c) C

The numbers on the X axes in the histograms are missing.

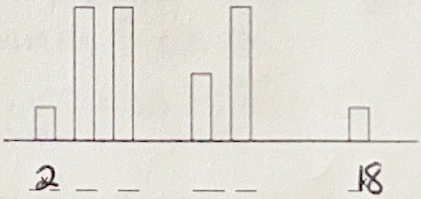
Histogram A



Histogram B



Histogram C



23. What number belongs in the middle of Histogram A? a) 4 b) 100 (c) 400 d) impossible to tell

24. What numbers belong under each of the 3 bars in Histogram B?
a) 1, 2, 5 (b) 1, 3, 9 c) 1, 2, 100 d) impossible to tell

25. What numbers belong under the first and last bar on Histogram C (marked with x's)?
a) 1, 100 (b) 2, 18 c) 1, 9 d) impossible to tell

26. Looking at the same box and drawing 100 times, what is the EV of the average from that box?
(a) 4 b) 100 c) 300 d) 400

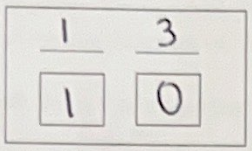
27. Looking at the same box and drawing 100 times, what is the SE of the average from that box?
(a) 0.3 b) 3 c) 30 d) 300

$\frac{SD}{\sqrt{n}} = 3/\sqrt{100} = 0.3$

28. There's a 95% chance that the average of the100 draws will be between _____ and _____.
a) 3.7, 4.3 (b) 3.4, 4.6 c) 370, 430 d) 340, 460

95% CI = $4 \pm 2(0.3)$

Now suppose you draw at random with replacement from the same box, but this time you're only interested in the percent of 9's.



Here is the original box: 1 3 3 9

29. To complete questions 30 and 31, you'll need to draw a new box. Draw the box using the figure to the left. How many 1s are in that box?
a) 0 (b) 1 c) 2 d) 3 e) 4

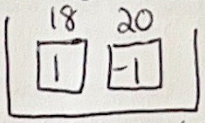
30. What is the EV of the percent of 9's in 100 draws?
a) 9% b) 20% (c) 25% d) 50% e) 75%

31. What is the SE of the percent of 9's in 100 draws?
a) 0.43% b) 0.5% c) 3.7% (d) 4.33% e) 5%

$$SE\% = \frac{SD}{\sqrt{n}} \times 100 = \frac{|1-0|\sqrt{1/4 \times 3/4}}{\sqrt{100}} \times 100 = 4.33$$

Questions 32-37 pertains to the following situation: 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 of the corresponding box is -0.0526, and the SD of the box is 1. Imagine playing 100 times.

32. The amount of money you get from playing this bet 100 times is like drawing from what box?
a) It has two tickets: 1 marked "1" and 1 marked "-1".
b) It has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
(c) It has 38 tickets: eighteen are 1's and twenty are -1's
d) It has 400 tickets: half are -1's, half are 1's.
e) It has 38 tickets: eighteen 1's, eighteen -1's, and two 0's



38 tickets

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

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

33. With or without replacement? (a) with b) without

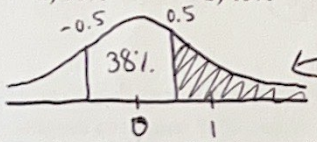
34. The expected value of the sum of the draws is closest to
a) $5.26 b) $-0.526 c) 0 d) $0.526 (e) $ -5.26

$$EVsum = n \times avg = 100 \times -\frac{2}{38} = -5.26$$

35. The standard error of the sum of the draws is closest to
a) $1 b) $ 5.26 (c) $10 d) $0.10 e) $100

$$SEsum = \sqrt{n} \times SD = \sqrt{100} \times 10$$

36. Use the normal curve to estimate the chance that you would win more than $0. The answer is closest to…
a) 10% b) 15% c) 20% (d) 30% e) 40%



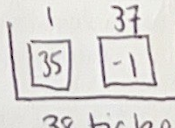
$$\frac{100-38}{2} = 31\%$$

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

37. If you play the game 500 times, will the chance that win more than $0 be larger than, smaller than, or the same as when you play 100 times? (a) smaller than b) larger than c) the same as

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.

38. What is the average of the box?
a) 0 b) 18 c) 0.89 d) -30/38 (e) 2/38



38 tickets

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

39. What is the SD of the box?
a) 0.5 b) 0.16 c) 1 d) 5.44 (e) 5.76

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

Questions 40-42 pertain to this scenario: A recent Gallup poll asked a random sample of 5,167 adults nationwide the following question: "Do you think your life will ever get completely back to the "normal" that existed before the coronavirus pandemic?" 47% of the sample answered "NO".

40. The EV of the percent of people in the nation would answer NO to this question is?
(a) 47% b) 53% c) 50% d) 0.47%

41. What is the SE%? a) 0.499% b) 35.88% (c) 0.69% d) Impossible to compute a SE for this sample.

$SE\% = |1-0|\sqrt{.47 \times .53}/\sqrt{5167} \times 100 = 0.69\%$.

42. Suppose I asked the same survey in class as an iClicker and got the exact same percentage of people saying no (47%)! What is the SE%? a) 0.499% b) 35.88% c) 0.69% (d) Impossible to compute a SE for this sample.

Question 43: The Census Bureau is planning to take a simple random sample amounting to 0.5% 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 = 500,000). a) quite a bit lower than b) about the same as (c) quite a bit higher than

Question 44: 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: (a) 1000 b) 2000 c) 4000 d) 8000 ↘ same sample size

Questions 45-49 pertain to the following situation: A recent poll conducted here in the US, asked a simple random sample of 1060 college professors whether they thought that their university should ban the use of ChatGPT. 35% of the sample favored a ban on ChatGPT.

45. What most closely resembles the relevant box model?
a) It has 1060 tickets, 35% are marked "1" and 65% are marked "0"
b) It has 1060 tickets with an average of 0.
(c) It has millions of tickets marked "0" and "1", but the exact percentage of each is unknown.
d) It has millions of tickets, exactly 35% are marked "1" and 65% are marked "0"

46. The draws are made _____replacement. a) With (b) Without

47. The SE of the sample percent is about 1.5% An approximate 95% confidence interval for the percentage of all US college professors who favor banning ChatGPT is:
a) (33.5%-36.5%) (b) (32%-38%) c) (30%-40%) d) (32.7%-35.3%) e) (92%-98%)

$95\%\ CI = 35\% \pm 2(1.5\%)$

48. Suppose 80 pollsters each randomly sampled 1060 US college professors asking them whether their university should ban ChatGPT. All 80 pollsters computed **90%** confidence intervals to estimate the percentage of all US college professors who think their university should ban ChatGPT. About how many of the 80 confidence intervals would **miss** the true population percentage?
a) 72 b) 20 c) 80 d) 106 (e) 8

$80 \times 0.9 = 72$ ← will contain true avg
$80 - 72 = 8$ ← will miss

49. If the sample size of the poll was increased by a factor of 4 (to n = 4240) then the width of the 90% confidence interval would…
a) increase by a factor of 2 b) increase by a factor of 4 (c) decrease by a factor of 2 d) decrease by a factor of 4

Questions 50-53 pertain to the following situation: 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 2024 presidential election.

50. About how many people would they have to poll to get a 95% Confidence Interval with a Margin of Error of 4%? Assume the SD=0.5. a) 400 (b) 625 c) 1111 d) 2500 e) 10,000

$$n = \left(\frac{100 \times 2 \times 0.5}{4}\right)^2 = 625$$

51. If we were to create a box model to model this situation, how many tickets would be in the box?
(a) 400,000 b) 625 c) millions d) impossible to tell

52. If we were to create a box model to model this situation, would we know the exact percentage of adults who will vote in the 2024 presidential election? a) Yes (b) No

53. If the city's population was 100,000 instead of 400,000 how should the survey organization adjust the sample size to keep the same margin of error?
a) Increase it by a factor of 2 b) Decrease it by a factor of 2 c) Increase it by a factor of 4
d) Decrease it by a factor of 4 (e) Keep it the same

Questions 54-60 pertain to the following survey:
As part of a survey on caffeine consumption among students at UIUC, a survey was given to a simple random sample of 400 UIUC students to see how much coffee they were drinking each week. In the survey, the students reported an average of 7.3 cups of coffee per week with an SD of 9.8 cups of coffee. Assume there are 35,000 undergraduates at UIUC.

54. What most closely resembles the relevant box model (the box from which the tickets are drawn)?
a) It has 35,000 tickets, 7.3% are marked "1" and 92.7% are marked "0"
b) It has 400 tickets marked with numbers, but the exact average and SD are unknown
(c) It has 35,000 tickets marked with numbers, but the exact average and SD are unknown.
d) It has 35,000 tickets marked with numbers and the average of the tickets is 7.3 and the SD is 9.8.

55. The draws are made _____replacement. a) With (b) Without

56. We can estimate the average number of cups of coffee among all UIUC students to be 7.3 cups per week, and the SE for the sample average is closest to? a) 9.8 (b) 0.5 c) 1 d) 0.1 $SE_{avg} = 9.8/\sqrt{400} = 0.49$

57. The best way to interpret a 95% confidence interval for the average number of cups of coffee UIUC students drink per week is:
a) About 95% of the students in the survey reported drinking between 7.3 ±1 cups of coffee per week.
b) About 95% of UIUC students have between 7.3 ±1 cups of coffee per week.
(c) 7.3 ±1 is a 95% confidence interval for the average number of cups of coffee all UIUC students drink per week.

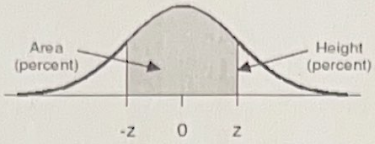
58. The 95% confidence interval in question 57 can be applied to:
a) All college students in the US
b) All Stat 100 students at UIUC
(c) All UIUC students
d) Both b and c are correct

59. In general, we can create confidence intervals for…
a) Any type of sample, as long as the sample size is large.
(b) Random samples only.
c) All samples involving surveys or public opinion polls.
d) All samples where the participants are asked yes or no questions.

60. If we wanted to create an 80% confidence interval, instead of a 95% confidence interval, how many standard errors should we add and subtract from the sample average? a) 1 b) 2 c) 1.5 (d) 1.3 e) impossible to tell

Exam 3 Formulas
EVsum = n*average of box SEsum = sqrt(n) * SD of box EVavg = average of box
SEavg = SD of box / sqrt(n) EV% = percent in box SE% = [SD of box / sqrt(n)] * 100%
Z = (Value – EV) /SE SD Shortcut Formula = |a – b|*sqrt(fraction of "a" tickets * fraction of "b" tickets)
n= (100*z*SD/Margin of Error)^2

STANDARD NORMAL TABLE



Standard Units

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