EXAM 3: Statistics 100

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)** L1 (Tues/Thurs at 12:30 pm) **ii)** KF (Flanagan Online) **iii)** ONL (Yu Online)

This test is ALL multiple choice. Circle all answers on this exam and fill in the corresponding bubble on your 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 Form 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 Student ID number (UIN) in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box.
- *No need to bubble in anything for Section or Form*.

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.

Make sure you have all 7 pages including the normal table (65 questions).

There is NO CLASS on Thursday!

Scores will be posted on Compass by Friday evening and exams will be returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours next week.

Questions 1-3 pertain to the situation below.
A 25 question multiple-choice test awards 4 points for each correct answer and subtracts 1 point for each incorrect answer. Each question has 4 choices only 1 of which is correct. Suppose you guess at random on each question and your score is computed.

1. Which is the appropriate box model?
a) The box has 4 tickets: 1 marked "1" and 3 marked "0"
b) The box has 25 tickets: 4 marked "1" and the rest marked "0"
c) The box has 4 tickets: 1 marked "4", and 3 marked "-1"
d) The box has 25 tickets, all marked either "4" or "-1", but the exact percentages of each type are unknown

2. Your score corresponds to the sum of _____draws from the box. **a)** 1 **b)** 4 **c)** 25 **d)** 100 **e)** none of these

3. These draws are taken _____replacement. **a)** with **b)** without

Questions 4-6 pertain to the situation below.
A gambler plays roulette 100 times. 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.

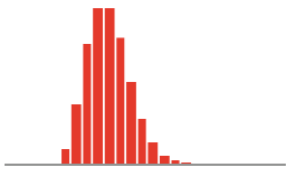
4. What is the appropriate box model?
a) The box has 100 tickets: 50 marked "1" and 50 marked "-1"
b) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
c) The box has 38 tickets: 18 marked "1", 18 marked "-1" and 2 marked "0"
d) The box has 38 tickets: 18 marked "1" and 20 marked "-1"
e) The box has 38 tickets: 18 marked "1" and 20 marked "0"

5. The gambler's winnings correspond to the sum of _____draws from the box.
a) 1 **b)** 18 **c)** 38 **d)** 100 **e)** None of these

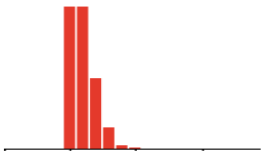
6. The draws are made _____ replacement. **a)** With **b)** Without

Questions 7-9 pertain to the situation below.
The 3 histograms below (in scrambled order) are the probability histograms for the sum of **23, 79 and 146** random draws with replacement from a box that has 49 tickets marked "0" and only 1 marked "1". Below each histogram, choose how many draws it has. Use each answer only once.

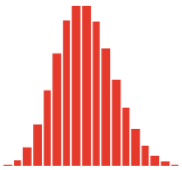
Histogram A



Histogram B



Histogram C



7. Histogram A corresponds to _____ draws.
a) 23 **b)** 79 **c)** 146

8. Histogram B corresponds to _____ draws.
a) 23 **b)** 79 **c)** 146

9. Histogram C corresponds to _____ draws.
a) 23 **b)** 79 **c)** 146

Questions 10-13 pertain to tossing a fair coin:
A coin is tossed 100 times and EVsum = 50 heads and SEsum = 5 heads and the EV%=50% and SE%=5%. Now suppose you toss the coin 3600 times.

10. The EVsum = ____ heads **a)** 6 **b)** 36 **c)** 50 **d)** 1800 **e)** 3600

11. SEsum = ___ heads **a)** 0.83 **b)** 5 **c)** 30 **d)** 36 **e)** 300

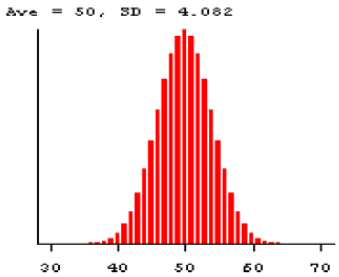
12. EV%= ___% **a)** 6 **b)** 36 **c)** 50 **d)** 1800 **e)** 3600

13. SE%=_____% **a)** 0.83 **b)** 5 **c)** 30 **d)** 36 **e)** 300

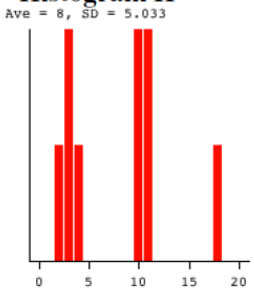
Questions 14 -17 pertain to the situation below.
The 4 histograms below (in scrambled order) are the probability histograms for the sum of 2 draws from Box A, 2 draws from Box B, 25 draws from Box A and 25 draws from Box B. Which histogram is which?

Box A: 1 2 3 Box B: 1 2 9

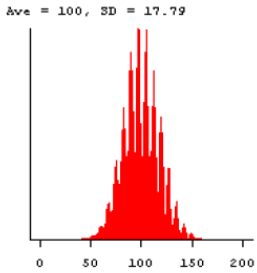
Histogram I



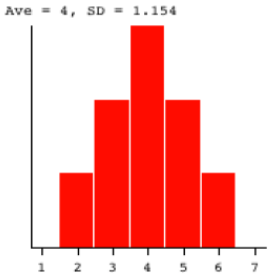
Histogram II



Histogram III



Histogram IV



14. Histogram I corresponds to _____draws from Box ____.
a) 2; A **b)** 25; A **c)** 2; B **d)** 25; B

15. Histogram II corresponds to _____draws from Box ____.
a) 2; A **b)** 25; A **c)** 2; B **d)** 25; B

16. Histogram III corresponds to _____draws from Box ____.
a) 2; A **b)** 25; A **c)** 2; B **d)** 25; B

17. Histogram IV corresponds to _____draws from Box ____.
a) 2; A **b)** 25; A **c)** 2; B **d)** 25; B

Questions 18-21 pertain to the following situation.
100 draws are made at random with replacement from the box containing 4 tickets: 1 2 4 5

18. The ***smallest*** the sum of the 100 draws could possibly be is:
a) 1 **b)** 5 **c)** 100 **d)** 300 **e)** None of these

19. The ***largest*** the sum of the 100 draws could possibly be is:
a) 5 **b)** 100 **c)** 300 **d)** 500 **e)** None of these

20. What is the **EV** (expected value) of the **sum** of the **100** draws?
a) 3 **b)** 5 **c)** 12 **d)** 300 **e)** 1200

21. What is the **SE** (Standard Error) of the sum of the 100 draws? (SD of box = 1.6)
a) 0.016 **b)** 0.16 **c)** 1.6 **d)** 5.06 **e)** 16

For the following questions (22-27), convert the same box (shown to the right) to a 0-1 box to count the amount of even numbers. Use this new box to answer questions 22-27. 1 2 4 5

22. This box would have _____ 1's and _____ 0's. Fill in the blanks with the amount of 1s and 0s.
a) 2, 2 **b)** 0, 1 **c)** 1, 3 **d)** 2, 4 **e)** None of these

23. How many **even** numbers (total) would you expect to get if you drew from this box 100 times?
a) 2 **b)** 50 **c)** 100 **d)** 200 **e)** None of these

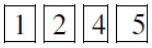
24. What is the **SD** of the **box** corresponding to the total number of evens you expect to get if you drew from this box 100 times?
a) 1.6 **b)** 0.05 **c)** 0.25 **d)** 0.5 **e)** 5

25. What is the **SE** of the total number of evens you expect to get if you drew from this box 100 times?
a) SD/$\sqrt{4}$ **b)** SD/$\sqrt{100}$ **c)** SD * $\sqrt{4}$ **d)** [SD/$\sqrt{100}$] * 100 **e)** SD *$\sqrt{100}$

26. What is the **percentage** of **evens** you would expect to get if you drew from this box 100 times?
a) 2% **b)** 50% **c)** 100% **d)** Not enough information **e)** None of these

27. What is the **SE** of the **percentage** of evens you expect to get if you drew from this box 100 times?
a) [SD/$\sqrt{4}$] *100 **b)** SD/$\sqrt{100}$ **c)** SD/$\sqrt{4}$ **d)** [SD/$\sqrt{100}$] * 100 **e)** SD *$\sqrt{100}$

For the same box (shown to the right), the expected value for the percent of 1's in 100 draws is EV=25% and the standard error is about SE=4%. Use this information to answer the following questions.



28. If I wanted to calculate the probability that the percentage of 1s in 100 draws is less than 30%, the corresponding z-score would be?
a) 1.25 **b)** -1.25 **c)** 1 **d)** -1 **e)** 1.2

29. If I wanted to calculate the probability that the percentage of 1s in 100 draws is less than 30%, the chance would be? Hint: Draw a normal curve.
a) 15% **b)** 21% **c)** 10.5% **d)** 79% **e)** 89.5%

Questions 30-35 pertain to the following situation.
A gambler plays roulette **81** times betting $1 on the numbers 1, 2 ,3 and 4 each time. If the ball lands on 1, 2 ,3 or 4, the gambler **wins $8**, if the ball lands on any of the other 34 numbers the gambler **loses $1**.
The roulette wheel has 38 slots numbered 1-36, 0 and 00.

30. Which is the appropriate box model?
a) The box has 38 tickets: 1 marked "1", 1 marked "2", 1 marked "3", 1 marked "4" and 34 marked "-1"
b) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
c) The box has 38 tickets: 4 marked "8" and 34 marked "0"
d) The box has 38 tickets: 4 marked "8" and 34 marked "-1"
e) The box has 81 tickets: 4 marked "8" and the rest marked "-1"

31. This score corresponds to the _____ of _____ draws from the box. Fill in the first blank with sum, average, or percent and the second blank with the number of draws.
a) sum; 81 **b)** average; 81 **c)** sum; 4 **d)** average; 38 **e)** percent; 38

32. What is the average of the box?
a) $\frac{1+2+3+4}{4}$ **b)** $\frac{1+2+3+4}{81}$ **c)** $\frac{4(8)+34(-1)}{2}$ **d)** $\frac{4(8)+34(-1)}{81}$ **e)** $\frac{4(8)+34(-1)}{38}$

33. What is the SD of the box?
a) 2.14 **b)** 0.5 **c)** 0.307 **d)** 9 **e)** 2.76

34. Use the normal approximation and the fact that the **EV Sum is about $ - 4** and the **SE Sum is about $25** to figure the chance that the gambler will win more than **$6** in 81 plays. What is the **Z score**?
a) 0.4 **b)** -3 **c)** 1 **d)** 0.08 **e)** None of these

35. Use the normal approximation and the fact that the **EV Sum is about $ - 4** and the **SE Sum is about $25** to figure the chance that the gambler will win more than **$6** in 81 plays. What is the **chance**? Hint: Draw a normal curve.
a) 31% **b)** 38% **c)** 69% **d)** 34.5% **e)** 16%

Questions 36-38 pertain to the following situation.
In November 2010, the week before California voters were to decide on Proposition 19 on whether to legalize marijuana, CNN/Time conducted a state-wide pre-election poll, asking a randomly selected sample of 888 likely California voters: "Will you oppose or support Proposition 19 to legalize marijuana in California?" 53% of the sample said they would vote to oppose the Proposition.

36. What is the SE of the sample %?
a) 0.499 **b)** 0.0167 **c)** 1.675 **d)** 0.835 **e)** Not possible to calculate

37. How did the election turn out? 54% of the nearly 8 million voters opposed the proposition. Which of the following statements best describes how good a job the CNN random poll did in predicting the election results?
a) The CNN poll was quite accurate in predicting the election since 54% is well within its 95% Confidence Interval.
b) The CNN poll was off by 1% of 8 million = 80,000 people which is a huge error considering that it's 90 times bigger than their sample size and thus outside of their 95% confidence interval**.**
c) It's impossible to determine the accuracy of the CNN poll because those who watch CNN are not typical voters.

38. A similar poll was asked to a random sample of students at the U of I in 2019: Do you support the legalization of marijuana? 66% of the students said yes and the SE% was 5%. What is an 85% confidence interval for the percent of students who say yes?
a) 85% ± 1*5% **b)** 66% ± 1.45*5% **c)** 85% ± 1.45*5% **d)** 66% ± 1*5%

Questions 39-43 pertain to the following situation:
This October, 4 polls asked this question: "Do you approve or disapprove of the impeachment charges against Donald Trump?" The Polling Report poll asked the question of a randomly selected sample of 1,650 adults nationwide. The MSNBC Live and Fox News Live polls simply posted the question on their websites and allowed anyone who visited their website to cast a vote, and I asked the question on our Bonus Survey 3. Here are the results:

	Disapprove	Approve	Sample Size
Polling Report	44%	56%	908
MSNBC Live	40%	60%	7,328
Fox News Live	97%	3%	8,421
Bonus Survey 3	34%	66%	770

39. Which poll best reflects how all US adults would answer this question?
a) The Polling Report because the sample was randomly selected from the entire US adult population.
b) The Fox Live Vote poll because it has the largest sample size.
c) The MSNBC Live Vote since it's likely to have the most informed respondents.
d) Bonus Survey 3 because the responses were anonymous.

For each poll listed below, is it possible to calculate a 95% Confidence Interval for the % of ALL US adults who would say they disapprove of? Choose "Yes" or "NO" for each poll.
40. Polling Report **a)** Yes **b)** No
41. Fox Live Poll **a)** Yes **b)** No
42. MSNBC Live Poll **a)** Yes **b)** No
43. Bonus Survey 3 **a)** Yes **b)** No

Questions 44-47 pertain to the following situation:
A CNN/Time Poll asked a random sample of 1,250 adults nationwide the following question: "Would you be willing to allow a reality-based television show film you drunk?" I asked the same question as an iClicker in class last semester. Here are the results of both surveys:

	Yes	No or Unsure	Sample Size
CNN/Time Poll	16%	84%	1,250
In Class iClicker Poll	35%	65%	601

44. If I wanted to create a box model for both polls, how many tickets would be in the CNN/Time poll box?
a) Millions **b)** 16 **c)** 1,250 **d)** impossible to create a box model for the CNN/Time poll

45. If I wanted to create a box model for both polls, how many tickets would be in the iClicker poll box?
a) Millions **b)** 35 **c)** 601 **d)** impossible to create a box model for the iClicker poll

46. 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 answer yes to this question and why?
a) The in class iClicker poll because we can be sure it was taken in real time and the results were anonymous to everyone in the class.
b) The CNN/Time survey because the people were randomly drawn from all adults nation-wide.
c) The CNN/Time survey because the sample size was larger.

47. What is SE of the sample percent for the in-class iClicker poll?
a) It's not possible to calculate a SE for this sample because it wasn't random.
b) It's not possible to calculate a SE for this sample because we don't know the size of the population.
c) The SE of the sample percent is approximately 1%
d) The SE of the sample percent is approximately 3%
e) The SE of the sample percent is approximately 5%

Questions 48-49. Pretend there was a 3rd poll that was randomly drawn from all US adults done by another polling agency. They took a random sample of 1,000 US adults and got 18% of them answering yes to the same question.

48. Then our best estimate for the percent of all US adults who would answer, "YES" to the above question is?
a) 18% **b)** 20% **c)** 82% **d)** 35% **e)** Impossible to answer

49. And the SE of our sample percent would be_____%
a) 0.38% **b)** 0.5% **c)** 12% **d)** 1.2% **e)** 3.8%

Questions 50-53 pertain to the following situation:
Suppose THE Ohio State University and University of Chicago both decide to do a poll of their undergraduates. Both universities want a margin of error of 4%. The undergrad population at Ohio State is about 9 times larger than the undergrad population at Chicago.

50. Other things being equal, the number of people you'd have to poll at Ohio State is _____ the number of people you'd have to poll at Chicago.
a) 9 times larger than **b)** 3 times larger than **c)** about the same as **d)** 3 times less than **e)** 9 times less than

51. How many people would you have to poll get a 95% Confidence Interval with a Margin of Error of 4% at Ohio State? (Assume SD=0.5)
a) 625 **b)** 6, 250,000 **c)** 10,000 **d)** 8, 836 **e)** none of the above

52. Suppose you wanted to double your accuracy (that is, decrease your margin of error from 4% to 2%). How should you adjust your sample size?
a) Increase it **b)** Decrease it **c)** Keep it about the same

53. How many people would you have to poll get a 95% Confidence Interval with a Margin of Error of 1% at Ohio State? (Assume SD=0.47)

a) 625 **b)** 6, 250,000 **c)** 10,000 **d)** 8, 836 **e)** none of the above

Questions 54-61 pertain to the following scenario:
A poll taken in Chicago, asked a random sample of 625 brides chosen from all brides who got married within the past 2 years in Chicago, how much money they spent on their weddings. The average was $72,000 with a SD of $25,000.

54. What most closely resembles the relevant box model?
a) It has thousands of tickets. The average of the tickets = $72,000 and the SD = $25,000
b) It has about thousands of tickets marked with "0"s and "1"s, but the exact percentage of each is unknown.
c) It has thousands of tickets. On each ticket is written a dollar amount. The exact average and SD are unknown but are estimated from the sample.
d) It has 625 tickets. The average of the tickets = $72,000 and the SD = $25,000
e) It has 625 tickets. On each ticket is written a dollar amount. The exact average and SD are unknown but are estimated from the sample.

55. What is the SE of the sample average?
a) 625 **b)** 625,000 **c)** 25,000 **d)** 2,880 **e)** 1,000

56. An approximate 95% confidence interval for the average amount all brides in Chicago who got married in the last 2 years would say they spent on their weddings is
a) $50,000-$100,000 **b)** $25,000-$125,000 **c)** $69,000-$71,000 **d)** $70,000-$74,000

To which of the following populations can we also apply the above 95% confidence interval?

57. All US brides who got married in the past 2 years **a)** Yes **b)** No
58. All teenage brides who got married in Chicago in the past 2 years **a)** Yes **b)** No
59. All brides who got married in Chicago in the past 25 years **a)** Yes **b)** No

60. Let's say another poll was done in New York City by Karle Flanagan and Albert Yu. They randomly sampled 5,625 new brides and asked the same question. How does this standard error compared to the standard error you calculated in Question 55?
a) It is the same **b)** It increases by a factor of 9 **c)** It increases by a factor of 3 **d)** It decreases by a factor of 9
e) It decreases by a factor of 3

61. If the study asked the brides what % of their income they had spent on the wedding, the relevant box model would contain tickets marked with…
a) Only 0s and 1s
b) Numbers ranging from 0 to 100
c) Numbers ranging from 0 to infinity.

Questions 62 and 63

62. What is the largest the SD of a 0-1 box can be? (A 0-1 box is a box that has only 0's and 1's)
a) 0 **b)** 0.5 **c)** 1 **d)** 100 **e)** infinity

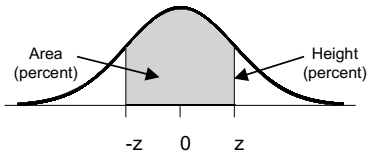
63. The SD of a 0-1 box is largest when the box contains _____% 0's and _____ % 1's. Fill in the two blanks with numbers.
a) 0, 100 **b)** 10, 90 **c)** 20, 80 **d)** 49, 51 **e)** 50, 50

Questions 64 and 65. A box has 35 tickets: 7 are marked -2 and 28 are marked 6. Hint: Draw the box first!

64. What is the average of the box?
a) 4.4 **b)** 154 **c)** 77 **d)** 2 **e)** 5.2

65. What is the SD of the box?
a) 0 **b)** 0.4 **c)** 1 **d)** 8.4 **e)** 3.2

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