**Form B Key**

**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)? _____

3) Which **section** are you in?

*Circle one:*      i) L1 (Karle Flanagan In Person)        iii) ONL (Karle Flanagan Online)

---

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 Form Directions**

- Print and bubble in your last name with **no spaces** in the YOUR LAST NAME box.
- Print and bubble in the first letter of your first name in the FIRST NAME INI box.
- Print and bubble in your University Identification Number (UIN) in the STUDENT NUMBER box.
- Print and bubble in your netID with **no spaces** in the NETWORK ID box.
- You do **NOT** need to fill out the SECTION bubbles or TEST FORM bubble parts of the scantron!
- *Write Stat 100 on the COURSE line.*
- *Write your instructor's name on the INSTRUCTOR line.*
- *Write your section on the SECTION line. Either L1 (in person) or ONL (online).*
- *Sign your name on the STUDENT SIGNATURE line.*

**READ THIS: Failure to fill out your scantron correctly will actually result in a loss of 2 points on your exam! At this point, everyone should be doing this correctly!!!!**

---

**WARNING-** The exams look alike but you are sitting next to people who 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 anything else you aren't allowed to have will result in a 0 and an academic integrity violation on your University record. By filling out this page of the exam, you are agreeing to follow the rules.

---

**Make sure you have all 8 pages including the normal table (64 questions).**

**There is NO CLASS on Friday!**

**Scores will be posted on Canvas by Monday at noon and exams will be returned in class next week. Online students may pick up their exam in 0060 Siebel Center for Design during office hours next week.**

Form B Key

**Questions 1-9 pertain to the following scenario:**
25 draws are made at random with replacement from the box containing these 5 tickets: $3, 4, 5, 6, 6$.

1. The smallest the sum of the 25 draws could possibly be is _____?
   a) 25          b) 50          **c) 75**          d) 120

2. The largest the sum of the 25 draws could possibly be is _____?
   **a) 150**          b) 120          c) 600          d) 300

3. What is the expected value (EV) for the sum of the draws?
   a) 600          b) 24          c) 100          **d) 120**

   $EV_{sum} = n \times avg = 25 \times 4.8 = 120$

4. What is the standard error (SE) for the sum of the draws? *The SD of the box is 1.2.*
   a) 0.048          **b) 6**          c) 30          d) 0.24

   $SE_{sum} = \sqrt{n} \times SD = \sqrt{25} \times 1.2 = 6$

Now suppose you draw at random with replacement from the same box above, but this time you're only interested in the percent of 6's that you get. What is the EV and SE of the percent of 6's in 25 draws? *Hint: draw a new box.*

5. Your new box should have _____ ticket(s) marked 1.
   a) 1          **b) 2**          c) 3          d) 4

   box: $\boxed{1}\;\boxed{1}\;\boxed{0}\;\boxed{0}\;\boxed{0}$   — 2, 3

6. Your new box should have _____ ticket(s) marked 0.
   a) 1          b) 2          **c) 3**          d) 4

   5 tickets

7. What is the expected value of the percent of 6's in 25 draws?
   a) 0.4%          b) 8%          c) 20%          **d) 40%**          e) 50%

8. What is the SD of your new box?
   a) 0.4          b) 0.24          **c) 0.49**          d) 0

   $SD = |1-0| \sqrt{\frac{2}{5} \times \frac{3}{5}} = 0.49$

9. What is the SE for the percent of 6's in 25 draws?
   a) 0.098%          b) 1.96%          c) 2.45%          d) 8%          **e) 9.8%**
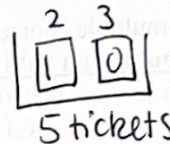
   $SE_{\%} = \frac{SD}{\sqrt{n}} \times 100$
   $= \frac{0.49}{\sqrt{25}} \times 100 = 9.8\%$

**For questions 10-15, fill in the blanks to make the chances equally likely in both scenarios:**
First, let's look at flipping a coin 56 times vs. 1,400 times.

10. The number of tosses (n) is increasing by a factor of _____?
    a) 0.04          **b) 25**          c) 1344          d) 5

    $\frac{1400}{56} = 25$

$28 \pm 4$ heads in 56 tosses is about as likely as _____ $\pm$ _____ heads in 1,400 tosses. *Fill in the first blank with the new EV and the second blank with the new SE that will make the chances equal.*

11. The first blank, or the new EV, is _____.
    a) 140          b) 50          c) 28          **d) 700**   half heads

12. The second blank, or the new SE, is _____.
    a) 0.16          **b) 20**          c) 0.8          d) 100

Next, we are looking at flipping a coin 100 times vs. 4,900 times.

13. The number of tosses (n) is increasing by a factor of _____?
    a) 0.02          **b) 49**          c) 7          d) 4,800

$50\% \pm 8\%$ heads in 100 tosses is about as likely as _____% $\pm$ _____% heads in 4,900 tosses. *Fill in the blank with the new EV and the second blank with the new SE that will make the chances equal.*
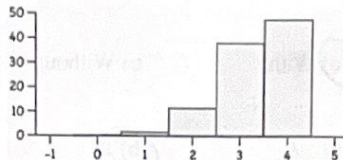
14. The first blank, or the new EV, is _____.
    a) 2,450          b) 7.14          c) 350          **d) 50**   50% heads

15. The second blank, or the new SE, is _____.
    **a) 1.14**          b) 56          c) 0.16          d) 8

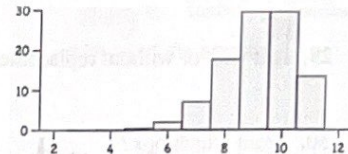**Questions 16-18 pertain to the follow situation:**
The 3 histograms below (in scrambled order) are the probability histograms for the sum of 4, 11, and 21 random draws with replacement from a box that has 12 tickets, 2 marked "0" and 10 marked "1".
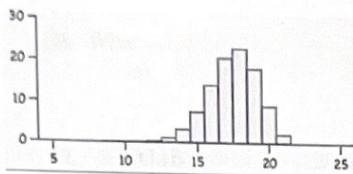


Histogram 1                          Histogram 2                          Histogram 3

16. Which histogram depicts 4 draws?     a) Histogram 1     b) Histogram 2 ⊙     c) Histogram 3

17. Which histogram depicts 11 draws?    a) Histogram 1     b) Histogram 2     c) Histogram 3 ⊙

18. Which histogram depicts 21 draws?    a) Histogram 1 ⊙     b) Histogram 2     c) Histogram 3

**The following situation pertains to questions 19-34:**
Fill in the first blank with the number of draws, the second blank with either "with" or "without", and the third blank with the letter corresponding to the appropriate box model. Choose from the box models below. Use each box model exactly once.

Box A
| 1 | -1/2 | -1/2 |

Box B
| 0 | 1 |

Box C
| 1 | 2 | 3 | 4 | 5 | 6 |

Box D
| 0 | 0 | 0 | 1 | 0 | 0 |

A die is rolled 125 times and the number of 1's is counted. This corresponds to drawing _____ times _____ replacement from Box ____.

19. How many times are we drawing?     a)   100     b) 1     c) 125 ⊙     d) 6

20. Is it with or without replacement?     a) With ⊙     b) Without

21. From which box?     a) A     b) B     ~~c)~~     d) D ⊙

A multiple-choice test has 32 questions. Each question had 3 options, only one of which is right. Suppose you randomly guess on all 32 questions. If you get a question right, you get 1 point, and if you get a question wrong, you lose half of a point. This corresponds to drawing _____ times _____ replacement from Box _____.

22. How many times are we drawing?     a)   1     b) 96     c) 3     d) 32 ⊙

23. Is it with or without replacement?     a) With ⊙     b) Without

24. From which box?     a) A ⊙     b) B     c) C     d) D

A fair die is rolled 8 times and the total number of spots is counted. This corresponds to drawing _____ times _____ replacement from Box ____.

25. How many times are we drawing?     a)   8 ⊙     b) 3     c) 6     d) 1

26. Is it with or without replacement?     a) With ⊙     b) Without

27. From which box?     a) A     b) B     c) C ⊙     d) D

A fair coin is tossed 46 times and the number of heads is counted. This corresponds to drawing _____ times _____ replacement from Box _____.

28. How many times are we drawing?  a)  2  (b) 46  c) 23  d) 100

29. Is it with or without replacement?  (a) With  b) Without

30. From which box?  a)  A  (b) B  c) C  d) D

_normal_  _lopsided_

Look at Box B and D above. The 4 histograms below are the probability histograms for the sum of 2 draws from Box B, 2 draws from Box D, 24 draws from Box B, and 24 draws from Box D. Which is which?



Histogram 1        Histogram 2        Histogram 3        Histogram 4
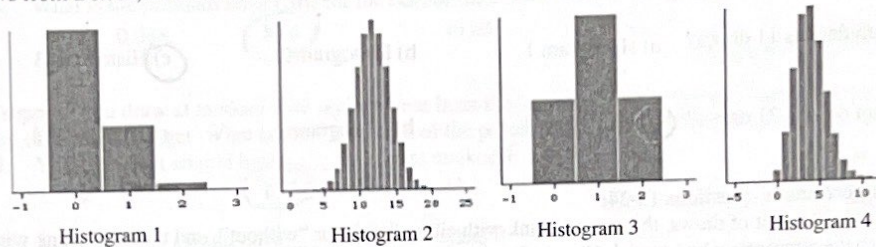
31. Histogram 1 represents how many draws and from which box?
    a)  2 draws from Box B
    (b)  2 draws from Box D
    c)  24 draws from Box B
    d)  24 draws from Box D

32. Histogram 2 represents how many draws and from which box?
    a)  2 draws from Box B
    b)  2 draws from Box D
    (c)  24 draws from Box B
    d)  24 draws from Box D

33. Histogram 3 represents how many draws and from which box?
    (a)  2 draws from Box B
    b)  2 draws from Box D
    c)  24 draws from Box B
    d)  24 draws from Box D

34. Histogram 4 represents how many draws and from which box?
    a)  2 draws from Box B
    b)  2 draws from Box D
    c)  24 draws from Box B
    (d)  24 draws from Box D

**The following scenario pertains to questions 35-42.**
A gambler plays roulette 49 times betting $1 on 8 numbers, 1-8, each time. If the ball lands on any of those 8 numbers, the gambler wins $4, if the ball lands on any of the other 30 numbers, the gambler loses $1. The roulette wheel has 38 slots numbered 1-36, 0, and 00.

35. Describe the appropriate box model.
    a)  The box has 49 tickets with 4 marked "1" and the rest marked "0"
    b)  The box has 49 tickets with 8 marked "4" and the rest marked "-1"
    c)  The box has 38 tickets with 8 marked "49" and the rest marked "0"
    (d)  The box has 38 tickets with 8 marked "4" and the rest marked "-1"
    e)  The box has millions of tickets with 8% of them marked "4" and the rest marked "-1"

8   30
4   -1
38 tickets

3
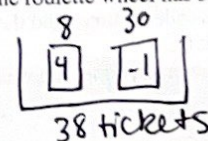
36. How many draws are made from this box?　(a) 49　b) 38　c) 30　d) 8

37. Are the draws made with or without replacement?　a) With　b) Without

38. What is the average of this box?
    a) 1/2　b) 4/38　(c) 2/38　d) 2/49　e) None of the above

$$avg = \frac{4(8) + 30(-1)}{38} = \frac{2}{38}$$

39. What is the SD of this box?
    a) 0.41　b) 0.5　c) 1.22　(d) 2.04　e) Impossible to calculate

$$SD = |4 - (-1)| \sqrt{\frac{8}{38} \times \frac{36}{38}} = 2.04$$

Use the normal approximation and estimate the EV to be about $-2 and the SE to be about $14 (approximately) to figure out the chance that the gambler will win less than $5 in 49 plays.

40. First, calculate the z-score
    a) 0.2　(b) 0.5　c) 0.8　d) 1　e) None of the above

$$z = \frac{5 - (-2)}{14} = 0.5$$

41. Now figure out the chance that the gambler will win less than $5 in 49 plays. I'd recommend drawing a normal curve, marking the z-score, and shading the area that represents winning less than $5. What is that chance closest to?
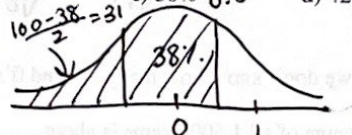    a) 21%　b) 31%　c) 38% 0.5　d) 42%　(e) 69%

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

38%

38% + 31% = 69%.

Suppose we were interested in how many times we'd expect the gambler to win playing 49 times (instead of how many dollars we'd expect him to win).

42. What is the appropriate box model?
    a) The box has 38 tickets: 8 marked "4" and 30 marked "0"
    b) The box has 38 tickets: one each of 1, 2, 3 … 36, 0, and 00
    (c) The box has 38 tickets: 8 marked "1" and 30 marked "0"
    d) The box has 38 tickets: 8 marked "1" and 30 marked "-1"

2　5　3
|1| |-7| |5|
10 tickets
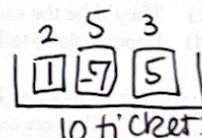
**The following situation pertains to questions 43-47.**
Suppose you randomly draw 64 marbles with replacement from a bag that contains 2 red marbles, 5 blue marbles, and 3 green marbles. If a red marble is drawn you win $1, if a blue marble is drawn you lose $7, if a green marble is drawn you win $5. Draw the appropriate box model by labeling the 3 tickets inside the box below with the correct numbers and writing how many of each ticket there are in the blanks above them. Assume the red marble tickets are on the far left, the blue marble tickets are in the middle, and the green marble tickets are on the far right. The tickets are the 3 rectangles.

43. What numbers go above the tickets?
    a) 1, -7, 5　(b) 2, 5, 3　c) 2, -35, 15　d) 81, 3, 10

44. What numbers go inside the tickets?
    (a) 1, -7, 5　b) 2, 5, 3　c) 2, -35, 15　d) 81, 3, 10

45. What is the average of the box?　a) -0.2　(b) 1.8　c) -9　d) -18　e) -31.5

46. What is the expected value of the sum of 64 draws (your net gain/loss)?
    a) -0.22　b) -14.4　c) -35.55　(d) -115.2　e) -145.8

$$avg = \frac{2(1) + 5(-7) + 3(5)}{10}$$
$$= -1.8$$

47. The SD of the box is $2.22. What is the SE of the sum of your winnings?
    a) -14.4　b) 0.28　c) -115.2　d) 142.08　(e) 17.76

$$EVsum = n \times avg = 64 \times -1.8 = -115.2$$
$$SEsum = \sqrt{n} \times SD = \sqrt{64} \times 2.22 = 17.76$$

4

**Questions 48-54 pertain to this situation:**
Last exam, I graded a random sample of 81 of the 1,500 exams to see how students would do. The average of the 81 exams was 83 with an SD of 10.

48. Which most closely resembles the relevant box model?
    a) The box has 1,500 tickets marked with "1"s and "0"s
    b) The box has 81 tickets marked with "1"s and "0"s. The exact percentages are unknown, but estimated from the sample
    c) The box has 81 tickets with an average of 83 and an unknown SD, estimated to be around 10
    d) The box has 1,500 tickets with an average of 83 and a SD of 10
    e) The box has 1,500 tickets, marked with numbers ranging from 0 to 100, the exact average and SD of the box are unknown and are estimated from the sample

49. How many draws from the box?  a) 1,500      b) 10      c) 83      d) 81      e) 64

50. Are the draws made with or without replacement?    a) With      b) Without      c) Either is fine

51. The best estimate for the average exams scores of all 1,500 students is _____.
    a) 10          b) 81          c) 83          d) Cannot be determined

52. What is the SE of the sample average?
    a) 1.11
    b) 0.12
    c) 0.9
    d) 1.25
    e) Unable to calculate because we don't know how many 1's and 0's are in the box.

$$SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{10}{\sqrt{81}} = 1.11$$

$68\%. \; CI = 83 \pm 1 SE$

53. A 68% confidence interval for the average of all 1,500 exams is about _____ to _____.
    a) 80.78, 85.22      b) 66.89, 69.11      c) 73, 93      d) 81.89, 84.11      e) 79.89, 82.11

54. Suppose we also computed a 95% CI for the average of all 1,500 exams. Which interval would be narrower?
    a) The 68% CI
    b) The 95% CI
    c) They'd be the same width
    d) Impossible to tell

**The following situation pertains to questions 55- 57.**
In October of 2020, 3 polls were taken asking who won the very last presidential debate: Joe Biden or Donald Trump. The Washington Times and Breitbart Polls posted the question on their websites, allowing anyone who visited to cast their vote. The CNN poll was based on the responses of 1,000 randomly selected adults nation-wide who watched the debate. Here are the results of each of the polls:

|                  | Biden | Trump | Sample Size |
|------------------|-------|-------|-------------|
| Washington Times | 18%   | 82%   | 25,000      |
| Breitbart Poll   | 29%   | 71%   | 80,000      |
| CNN              | 57%   | 43%   | 585         |

55. As you can see, the results of the 3 polls are quite different. Which poll gives the best estimate of the percentage of all U.S. adults who watched the debate who thought Donald Trump won?
    a) The Breitbart Poll because it has the largest sample size
    b) The Washington Times Poll because it gives the most definitive result of who won the debate
    c) The CNN Poll because the people were randomly chosen from all adult viewers nation-wide
    d) Both the Breitbart and CNN Poll can be trusted since they are very close to each other

56. Can we calculate the SE for the percent of the Washington Times Poll? Choose the best answer.
    a) Yes, the SE% is approximately 2.43%
    b) Yes, the SE% is approximately 0.3%
    c) No because the sample was too big
    d) No because this poll was not randomly selected

$EV_{sum} = n \times avg = 64 \times -1.8$

$SE_{sum} = \sqrt{n} \times SD = \sqrt{64} \times 2.99 = 1717.76$

Another poll was done by Gallup asking the same question as above. The results were from 1,200 <u>randomly sampled adults</u> <u>nationwide who watched the debate.</u>

57. If we computed a 95% confidence interval for the percentage of people who think Joe Biden won the debate, to which of the following populations can we apply that interval?
   a) All U.S. adults
   b) All U.S. adults who plan to vote
   c) All U.S. adults who watched the debate
   d) All Biden supporters
   e) Everyone in the world

## The following situation pertains to questions 58-60.
Suppose a government survey organization took a simple random sample of 1,000 people in Illinois and computed the SE% and a 95% confidence interval in order to estimate the percentage of all voters in Illinois who support making public colleges and universities tuition free.

58. If they decided to increase the sample size to 16,000, the new SE% would....
   a) Be divided by 4
   b) Be divided by 16
   c) Stay the same
   d) Be multiplied by 4
   e) Be multiplied by 16

*Law of Averages*
*as $n \uparrow$ by 16, SE% $\div \sqrt{16} = 4$*

59. If the sample size was increased to 16,000, the width of the new 95% confidence interval would....
   a) Be divided by 4
   b) Be divided by 16
   c) Stay the same
   d) Be multiplied by 4
   e) Be multiplied by 16

Suppose 100 pollsters each randomly sampled 1,000 Illinois voters asking the same question. All 100 pollsters computed 85% confidence intervals to estimate the percentage of all Illinois voters who support making public colleges and universities tuition free.

60. About how many of the 100 confidence intervals would **contain** the true population percentage?
   a) 0          b) 15          c) 85          d) 100          e) None of the above

## Questions 61 is related to the following situation:
Among Us is a population online game that Karle and the TAs like to play together on Fridays! In the game, you are either assigned to be an "imposter" or a "crewmate". Last week, we looked at 2 polls asking whether or not people prefer to be an "imposter" or "crewmate". The first poll was a Buzzfeed poll that was posted on their website, allowing anyone who visited to cast their vote. The second pol was posted on Karle's Instagram story, allowing anyone who viewed it to cast their vote. Here are the results of each of the polls:

|                       | Impostor | Crewmate | Sample Size |
|-----------------------|----------|----------|-------------|
| Buzzfeed Poll         | 61%      | 39%      | 108, 300    |
| Karle's Instagram Poll | 57%      | 43%      | 159         |

61. As you can see, the results of the 2 polls are quite different. Do these polls give the best estimate for all U.S. adults who play Among Us?
   a) The Buzzfeed Poll is the best because it has a much bigger sample size than the Instagram Poll
   b) The Instagram Poll is better because the people who took it are all adults
   c) Both the Buzzfeed and the Instagram Poll are best since the results are so close to each other
   d) Neither of these polls are best – taking a random sample of all U.S. adults who play Among Us would be best and neither of these are random samples

**Questions 62-64 pertain to the following scenario:**
In pre-election polls in close races, pollsters often want a small margin of error. These races are extremely hard to predict because all polls using samples have a certain amount of error. Let's say that back in 2020, we wanted to take a poll in both Nevada and Pennsylvania the day before the election to see who people planned on voting for. Both states wanted a small margin of error of 1%. The population in Pennsylvania is about 13 million and the population in Nevada is about 3 million.

62. Other things being equal, in order to obtain the same accuracy in the two polls, the number of people you'd have to poll in Pennsylvania is _____ the number of people you'd have to poll in Nevada.

    a) Smaller than
    b) Equal to *(circled)*
    c) Larger than
    d) Unable to answer because there was some fraudulent voting that occurred in Nevada

63. How many people would you have to poll to get a 95% confidence interval with a margin of error of 1% in Nevada? *Assume SD = 0.45.*

    a) 90        b) 9,216        c) 8,100 *(circled)*        d) 100        e) 10,000

$$n = \left( \frac{100 \times 2 \times 0.45}{1} \right)^2 = 8100$$

64. Say Nevada changed their minds and wanted an even smaller margin of error, only 0.5%. How should they adjust their sample size to get a smaller margin of error? (Still assume SD = 0.45).

    a) Decrease it        b) Increase it *(circled)*        c) Keep it the same

## 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%

SD Shortcut Formula = |a − b| * sqrt(fraction of "a" tickets * fraction of "b" tickets)

Z = (Value − EV) /SE