Exam 3 Statistics 100

EXAM 3: Statistics 100 Form B Answer Key Spring 2022

Read all of the directions below and make sure you fill out your scantron correctly! 54-accept a+b 55-accept a+b

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 below.

i) L2 (Karle Flanagan In Person) **ii)** O1 (Karle Flanagan Online) **iii)** O2 (Jonas Reger 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 Directions

- Print and bubble in your LAST NAME with **no spaces** starting in the left most column.
- Print and bubble in your FIRST INITIAL in the right-most column.
- Print and bubble in your University Identification Number (UIN) in the Student Number box.
- Print and bubble in your NET ID with **no spaces** in the NETWORK ID box (ex. kflan).
  - Be sure to include the numbers. Do not bubble in any dashes.
- *Write Stat 100* on the COURSE line.
- *Write your instructor's name* on the INSTRUCTOR line.
- *Write your section (L2, O1, or O2) on the SECTION line.*
- Sign your name, and right underneath the student signature line PRINT your name

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 8 pages including the normal table (70 questions).

There is NO CLASS on Thursday!

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.

Questions 1-12 pertain to the following situation: 25 draws are made at random with replacement from the box containing 4 tickets: [2] [4] [4] [10] avg = $\frac{2+4+4+10}{4} = 5$

1. The ***smallest*** the sum of the 25 draws could possibly be is? **a)** 8 **b)** 25 **c)** 2 **d)** 50 **e)** 20
2. The ***largest*** the sum of the 25 draws could possibly be is? **a)** 20 **b)** 35 **c)** 15 **d)** 40 **e)** 250
3. What is the **EV** (expected value) of the **sum** of the **25** draws? $EV_{sum} = n \times avg = 25 \times 5 = 125$
   a) 25 **b)** 35 **c)** 125 **d)** 5 **e)** 20
4. What is the **SE** (standard error) of the **sum** of the **25** draws? **(SD of box = 3)** $SE_{sum} = \sqrt{25} \times 3 = 15$
   a) 0.6 **b)** 15 **c)** 0.12 **d)** 75 **e)** 60
5. What is the **EV** of the **average** of the **25** draws? EV_{avg} = avg of box = 5
   a) 5 **b)** 20 **c)** 125 **d)** 0.2 **e)** 25
6. What is the **SE** of the **average** of the **25** draws? **(SD of box = 3)** $SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{3}{\sqrt{25}} = \frac{3}{5}$
   a) 0.6 **b)** 15 **c)** 60 **d)** 0.12 **e)** 3

Now suppose you draw at random with replacement from the same box above, but this time you're only interested in looking at the percent of 2's you get. What are the **EV** and the **SE** of the **percent** of 2's in **25** draws? ***(Hint: draw a new box)***

7. The new box should have…
   a) 25 tickets: 5 marked "1" and 20 marked "0"
   b) 4 tickets: 1 marked "1" and 3 marked "0"
   c) 25 tickets: 1 marked "2" and the rest marked "0"
   d) 4 tickets: 1 marked "2", 2 marked "4", and 1 marked "10"
   e) Millions of tickets: The exact percentage of 1s and 0s is unknown, but can be estimate from our sample

8. The EV of the percent of 2's in 25 draws is:	**a)** 6.25	**b)** 25	**c)** 125	**d)** 1	**e)** 50
9. What's the SD of the new box?	**a)** 1	**b)** 0.5	**c)** 3	**d)** 0.2	**e)** 0.433
10. What is the SE% of 2's in 25 draws?	**a)** 0.086	**b)** 1.732	**c)** 2.165	**d)** 8.66	**e)** 10

$SD = |1-0|\sqrt{\frac{1}{4} \times \frac{3}{4}} = 0.433$ $SE\% = 0.433/\sqrt{25} \times 100 = 8.66$

Next, use the normal approximation to find the chance that the sum of 400 draws will be above 1880? ***The EVsum= 2000 and the SEsum= 60 for 400 draws.***

$\frac{1880-2000}{60}$

11. First calculate the Z score. The z-score is? **a)** -2 **b)** 1 **c)** -1 **d)** -0.3 **e)** 2
12. Next, use the normal curve to figure out the chance that the sum of the 400 draws will be more than ~~1940~~ 1880. The chance is? **a)** 97.5% **b)** 95% **c)** 2.5% **d)** 5% **e)** 84%

$\frac{100-95}{2} = 2.5$ $Z = \frac{val - EV}{SE} = \frac{1880-2000}{60} = -2$

Chance = middle area + right tail = 95 + 2.5 = 97.5

Questions 13-18 pertain to the following situation: 1100 students will take this exam. Suppose tomorrow at the grading meeting I randomly pick 36 exams to grade to estimate the average of all 1100 exams. The sample of the **36** exams has an **average = 85 with a SD = 10.**

13. Which most closely resembles the relevant box model?
   a) The box has 36 tickets, marked with numbers ranging from about 20 to 100.
   b) The box has 36 tickets marked with "1"s and "0"s
   c) The box has 1100 tickets, marked with numbers ranging from about 20 to 100, the exact average is unknown but estimated from the sample.
   d) The box has 1100 tickets marked with "1"s and "0"s
   e) The box has 1100 tickets with an average of 85 and a SD of 10.

14. How many draws from the box? a) 1100 b) 10 c) 85 d) 36
15. The draws are made… a) with replacement b) without replacement
16. The best estimate for the average of all 1100 exams is? a) 36 b) 85 c) 10 d) Cannot be determined
17. What is the SE of the sample average? a) 1.25 b) 1.66 c) 0.27 d) 0.30 e) 60

$$SE_{avg} = \frac{SD}{\sqrt{n}} = 10/\sqrt{36} = 1.67$$

18. The exam scores do not follow the normal curve. Is it still possible to construct a 95% confidence interval for the average exam score of all 1100 students?
   a) Yes, even though the data doesn't follow the normal curve, the probability histogram for the average of 64 draws will come pretty close to following the normal curve.
   b) No, if the data does not follow the normal curve, it's never possible to construct confidence intervals
   c) No, it's not possible to construct confidence intervals for averages.

The next questions pertain to playing roulette: A gambler plays roulette **400** times betting \$1 on the numbers 21 and 29 each time. If the ball lands on either 21 or 29 the gambler wins \$17, if the ball lands on any of the other 36 numbers the gambler loses \$1. The roulette wheel has 38 slots numbered 1-36, 0 and 00.

19. Which is the appropriate box model? 2 winners
   a) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
   b) The box has 38 tickets: 2 marked "17" and 36 marked "-1"
   c) The box has 38 tickets: 1 marked "21" and 1 marked "29" and 36 marked "-1"
   d) The box has 38 tickets: 1 marked "17" and 37 marked "-1"
   e) The box has 50 tickets: 21 marked "17" and 29 marked "-1"

20. How many draws from the box? a) 2 b) 38 c) 17 d) 400
21. The draws are made… a) without replacement b) with replacement
22. What is the average of the box? a) -0.053 b) -0.526 c) 1.84 d) 33.05

$$avg = \frac{2(17)+36(-1)}{38} = -0.05$$

23. What is the SD of the box? a) 3.57 b) 0.22 c) 18 d) 4.02

72
7444

$$SD = |17-(-1)|\sqrt{2/38 \cdot 36/38} = 4.02$$

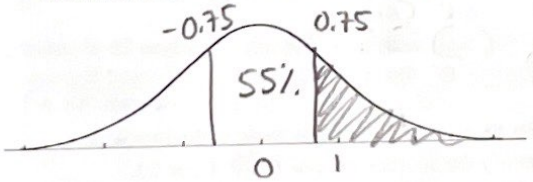
Use the normal approximation and the fact that the **EV = \$ - 21** and the **SE = \$80** (approximately) to figure the chance that the gambler will win more than **\$39** in 400 plays.

39--21
80

24. First calculate the Z score. a) -0.225 b) 0.75 c) 0.5 d) 0.225 e) -0.75

$$Z = \frac{val-EV}{SE} = \frac{39-(-21)}{80} = 0.75$$

25. Now mark the z score on the curve, shade the area representing the chance of winning **more** than \$39 in 400 plays.
   The chance is? a) 31% b) 55% c) 22.5% d) 45% e) 77.5%



$$\text{chance} = \text{right tail} = \frac{100-55}{2} = 22.5\%$$

Questions 26-43 pertain to the following situation: A slacker student named John Snow has 4 finals. Each final consists of 100 multiple-choice questions. He knows nothing (you know nothing, John Snow!) so he decides to randomly guess on every question so he can complete each final in less than 5 minutes.

Part 1: To compute the expected value (EV) for John's score for each final, you may need additional information. Which of the following do you need to know? Choose "Yes" if needed or "No" if not.

26. How much time is allotted for the exam. a) Yes b) No
27. How many points are awarded or deducted for each choice. a) Yes b) No
28. How many students are taking each final. a) Yes b) No
29. How many choices there are for each question. a) Yes b) No

30. Randomly guessing on all questions corresponds to drawing how many times from the appropriate box model?
a) 20 **b)** 4 **c)** 5 **d)** 100 **e)** millions

31. Do you draw with or without replacement from the appropriate box model? **a)** without **b)** with

Part 2: For questions 7-10 match the final exams to their corresponding box models. ***Use each box model exactly once***.

Box A: -1 1 **Box B:** 0 1 2 **Box C:** 0 1 **Box D:** -1 0 1

32. Final 1- Each question has 3 choices, one is a right answer, one is a wrong answer, and one is an "I don't know" answer. Your score is computed as the number of right answers minus the number of wrong answers. The "I don't know" answers are scored as 0 points.
This corresponds to Box… **a)** A **b)** B **c)** C **d)** D

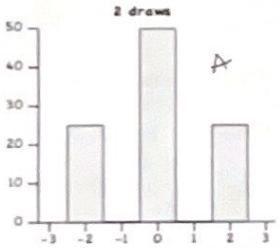
33. Final 2- Each question has 3 choices, one is the best answer and awarded 2 points, one is a mediocre answer and awarded 1 point, and one is a wrong answer and awarded no points.
This corresponds to Box… **a)** A **b)** B **c)** C **d)** D

34. Final 3- Each question is a true/false question. Your score is the number of answers you get right.
This corresponds to Box… **a)** A **b)** B **c)** C **d)** D

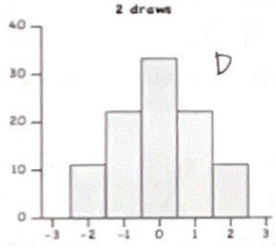
35. Final 4- Each question is a true/false question. Your score is the number of answers you get right minus the number of answers you get wrong.
This corresponds to Box… **a)** A **b)** B **c)** C **d)** D

Part 3: The 4 histograms below represent the probability histogram for the **sum of 2 draws** made at random with replacement from each of the boxes in part 2 above. For each histogram identify the appropriate box (A, B, C, or D).

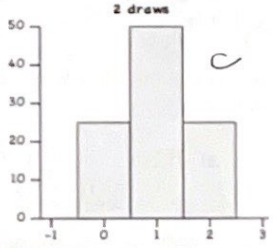
Histogram I



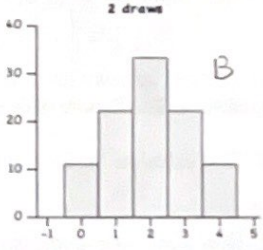
Histogram II



Histogram III



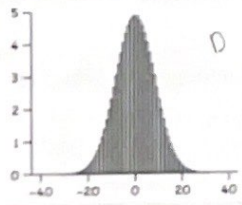
Histogram IV



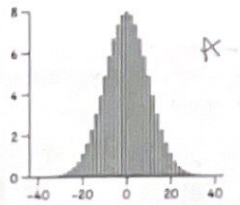
36. Histogram 1 represents 2 draws from Box…? **a)** A **b)** B **c)** C **d)** D
37. Histogram 2 represents 2 draws from Box…? **a)** A **b)** B **c)** C **d)** D
38. Histogram 3 represents 2 draws from Box…? **a)** A **b)** B **c)** C **d)** D
39. Histogram 4 represents 2 draws from Box…? **a)** A **b)** B **c)** C **d)** D

Part 4: The 4 histograms below represent the probability histogram for the **sum of 100 draws** made at random with replacement from each of the boxes in above. For each histogram identify the appropriate Box (A, B, C, or D).

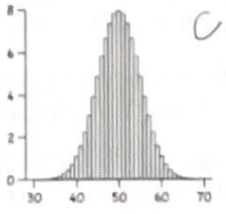
Histogram I
EV=0 SE=8.165



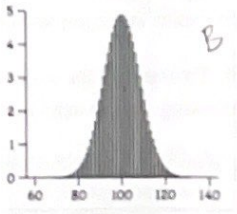
Histogram II
EV=0 SE=10



Histogram III
EV=50 SE=5



Histogram IV
EV=100 SE=8.165



HINT—The EVsum and SEsum of 100 draws are given above each histogram.

40. Histogram 1 represents 100 draws from Box…? **a)** A **b)** B **c)** C **d)** D

41. Histogram 2 represents 100 draws from Box…? **a)** A **b)** B **c)** C **d)** D

42. Histogram 3 represents 100 draws from Box…? **a)** A **b)** B **c)** C **d)** D

43. Histogram 4 represents 100 draws from Box…? **a)** A **b)** B **c)** C **d)** D

Questions 44-48 pertain to the following situation: The polling organization PPP conducted a nationwide Halloween poll in which they asked 1,111 **randomly** selected adults the following questions with these (rounded) results:

"Do you believe..?"	Believe	Don't Believe
in **Ghosts**?	40%	60%
that houses can be **Haunted**?	50%	50%
that people can become possessed by **Demons**?	60%	40%
that **Black Cats** can cause a change of luck?	70%	30%

44. The sample size for each question is the same. Is the SE of the sample percent the same for each of these questions?

a) No, the SE would be largest for belief in Haunted Houses and smallest for Black Cats changing one's luck.

b) No, the SE would be largest for belief in Black Cats changing one's luck and smallest for Haunted Houses.

c) Yes

45. Which question has a bigger SE—belief in Ghosts or belief in Demons, or are they the same?

a) Demons **b)** Ghosts **c)** They're exactly the same.

Fill in the blank below to construct the correct confidence interval.

46. An approximate **90%** Confidence Interval for the % of all US adults who would say they believe in **Ghosts** is

$40\% +/- \_\_\_\_\_ \times \frac{\sqrt{0.4 * 0.6}}{\sqrt{1111}} \times 100\%$ **a)** 1.65 **b)** 1 **c)** 2 **d)** 0.9 **e)** 0.4

47. If the researcher increased the sample size from 1111 to 4444 then the width of each confidence interval above would…. **a)** be multiplied by 2 **b)** be multiplied by 4 **c)** be divided by 2 **d)** be divided by 4 **e)** stay the same

48. Which of the following statements is correct?

a) If the researcher increased the sample size, we can't say how the standard error or accuracy would change.

b) If the researcher increased the sample size, the standard error would decrease and the accuracy would decrease.

c) If the researcher increased the sample size, the standard error would decrease and the accuracy would increase.

d) If the researcher increased the sample size, the standard error would increase and the accuracy would decrease.

e) If the researcher increased the sample size, the standard error would increase and the accuracy would increase.

Questions 49-52 pertain to the following situation. Fill in the following chart for the EV and SE of the number (sum) and the percent of heads in 100 tosses of a fair coin, the first row is done for you for 4 tosses:

n = # of tosses	**$EV_{\%}$**	**$SE_{\%}$**	**EV_{sum}**	**SE_{sum}**
4	**50%**	**25%**	**2**	**1**
100	Blank 1	Blank 2	Blank 3	Blank 4

49. What goes in Blank 1? **a)** 50 **b)** 2 **c)** 10 **d)** 0.5 **e)** 250

50. What goes in Blank 2? **a)** 1 **b)** 125 **c)** 6.25 **d)** 625 **e)** 5

51. What goes in Blank 3? **a)** 25 **b)** 10 **c)** 0.4 **d)** 50 **e)** 100

52. What goes in Blank 4? **a)** 0.04 **b)** 0.2 **c)** 5 **d)** 25 **e)** 50

The next few questions pertain to this situation: A nationwide Gallup poll asked a random sample of **1016** adults: "Are you afraid of public speaking in front of an audience?" **40%** of the people in the sample answered "YES".

53. What most closely resembles the relevant box model?

a) It has millions of tickets, exactly 40% are marked "1" and exactly 60% are marked "0"

b) It has millions of tickets marked "1" and "0", the exact percentage of each is unknown but are estimated from the sample to be 40% and 60% respectively.

c) It has 1016 tickets, 40% are marked "1" and 60% are marked "0"

54. The poll reported a Margin of Error = 3%. How did they get that number?

a) It's the SE of the sample percent **b)** It's 2 x SE of the sample percent. **c)** It's the SD of the sample

accept both

55. We can be about 95% confident that if we polled all US adults the percent who would say they afraid of public speaking in front of an audience would be between _____% and _____%.

a) 37%-43% **b)** 38.5%-41.5% **c)** 57%-63% **d)** 93%-97%

accept both 95% CI = sample % ± 2SE

Suppose 100 pollsters each randomly sampled 1,061 adults nationwide asking the same question. All 100 pollsters computed 85% confidence intervals to estimate the percentage of all US adults who would answer "Yes" to the question.

56. About how many of the 100 confidence intervals would miss the true population percentage?

a) 15 **b)** 50 **c)** 85 **d)** 100 **e)** Cannot determine from the information given

57. Let's say Gallup also asked a random sample of 1500 college students how many presentations they have given this past academic year. They find the sample mean and calculate that a 90% confidence interval (CI) for the average amount of presentations given this past academic year is 4 +/- 2. How do you interpret this 90% CI?

a) We are 90% sure that if we polled all college students, the true average would be between 2 and 6 presentations.

b) We are 90% sure that all college students have given between 2 and 6 presentations this past academic year.

c) 90% of all college students gave between 2 and 6 presentations this past academic year.

Questions 58-61 pertains to the following situation: During the same week, 3 polls asked the same question: "Do you think it is more important to protect gun rights or control gun violence?" The Marist poll asked that question of a ***randomly*** selected sample of 1,001 adults nationwide, Stat 100 students were asked that question on a Bonus Survey, and the WKRN.com poll simply posted the question on its website and allowed anyone who visited the website to cast a vote. Here are the results:

	Protect Gun Rights	Control Gun Violence	Sample Size
Marist Poll random	48%	52%	1,001
WKRN.com Poll	60%	40%	2,550
Bonus Survey	18%	82%	912

58. Which poll best reflects how all US adults would answer this question?

a) The WKRN Poll because it has the largest sample size.

b) The Marist random poll because the sample was ***randomly*** selected from the entire US adult population.

c) The Stat 100 Bonus Survey results because we know it was anonymous.

For each poll listed below, is it possible to calculate a 95% Confidence Interval for how all US adults would respond to the question?

59. Bonus Survey **a)** Yes or **b)** No

60. Marist Poll **a)** Yes or **b)** No

61. WKRN.com Poll **a)** Yes or **b)** No

The next few questions pertain to this situation and table: Last August, Pew Research asked a **randomly** selected sample of 2,231 US adults: "During the past 12 months, have you personally experienced discrimination or been treated unfairly because of your race or ethnic background, or not?" Here are the results.

	Yes, I have	No, I have not	Sample Size
All US Adults	22%	78%	2,231
White Adults	10%	90%	1,637
Black Adults	72%	28%	376
Latinx Adults	65%	35%	218

62. True or False? The reason a larger percentage of White adults answered "No" than Black adults (90% vs. 28%) is because the sample size was larger for the White adults (1,637 vs. 376). **a)** True **b)** False

63. What is the SE of the sample percent among Black adults?

a) The SE of the sample percent is about 2.3%
b) The SE of the sample percent is about 0.45%
c) The SE of the sample percent is about 0.89%
d) It's not possible to calculate a SE for this sample because we don't know the SD of the sample.
e) It's not possible to calculate a SE for this sample because we don't know the size of the population.

$$SE_{\%} = \frac{\sqrt{.72 \times .28}}{\sqrt{376}} \times 100 = 2.3\%$$

64. A 68% confidence interval for the percent of **all US adults** who would answer "yes, they have" is closest to:
a) 20.25% - 23.75% **b)** 21.12%- 22.88% **c)** 69.68-74.32% **d)** None of these are correct

$$SE_{\%} = \frac{\sqrt{.22 \times .78}}{\sqrt{2231}} \times 100 = 0.88$$

$$68\%\ CI = \text{sample } \% \pm 1\, SE_{\%} = 22 \pm 1(0.88)$$

To which of the following populations can we also apply the above 68% confidence interval from Question 19?

65. All Illinois adults **a)** Yes or **b)** No
66. All Black adults in the US **a)** Yes or **b)** No
67. All White adults in the US **a)** Yes or **b)** No

The last few questions pertain to the following situation: Suppose a survey organization is planning to take a random poll in Illinois (population about 9 million adults) and a random poll in US (population about 225 million adults) to estimate the percent of adults at both the state and the national level who would support legislation to ban assault weapons.

68. Other things being equal, to achieve the same level of accuracy in both polls, the number of people you'd have to poll in Illinois is about _____ the number of people you'd have to poll in the whole US.
a) 5 times larger than **b)** 25 times larger than **c)** 5 times smaller than **d)** 25 times smaller than **e)** the same as

69. How many people would you have to poll in the US to get a 95% Confidence Interval with a Margin of Error of **2%?** (Assume the SD of the population is close to 0.5)
a) 1111 **b)** 400 **c)** 2500 **d)** 10,000 **e)** 625

$$n = \left(\frac{100 \times z \times SD}{\text{M of E}}\right)^2 = \left(\frac{100 \times 2 \times .5}{2}\right)^2 = 2500$$

70. How many people would you have to poll in the Illinois to get a 95% Confidence Interval with a Margin of Error of **4%?** (Assume the SD of the population is close to 0.5)
a) 2500 **b)** 10,000 **c)** 625 **d)** 1111 **e)** 400

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

Exam 3 Formulas

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