EXAM 3: Statistics 100

READ THE DIRECTIONS BELOW TWICE!

Cover Sheet Questions

1) What's your **name**? KEY Form B _____

(Last name) (First name)

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

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

i) L1 (In Person Section) **ii)** ONL (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 (L1 or ONL) 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.

Make sure you have all 8 pages including the normal table (58 questions).

There is NO CLASS on Friday this week!

Scores will be posted on Canvas by Monday at 5pm. Students may pick up their exam in 171 Computer Applications Building during office hours next week.

LAW OF AVGS

Questions 1-6 pertain to the following situation: In tossing a fair coin, follow the steps to make the chances equally likely in both scenarios.

Scenario 1: 15 ± 3 heads in 30 tosses is about as likely as getting ____ ± _____ heads in 750 tosses.

1. Compare the number of tosses in both cases. The number of tosses (n) is increasing by a factor of_____
   a) 1 b) 15 c) 25 = n d) 30 e) 50 750/30 = 25
2. What goes in the first blank? In other words, what is the EV?
   a) 0 b) 15 c) 25 d) 75 e) 375 $\frac{1}{2}(750) = 375$
3. What goes in the second blank? In other words, what is the SE?
   a) 3 b) 15 c) 30 d) 50 e) 75

SE x $\sqrt{n}$ = 3 x $\sqrt{25}$ = 3 x 5 = 15

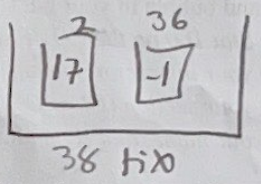
Scenario 2: 50% ± 6% heads in 65 tosses is about as likely as getting ____ ± ____ heads in 585 tosses.

4. Compare the number of tosses in both cases. The number of tosses (n) is increasing by a factor of_____
   a) 5 b) 9 c) 15 d) 65 e) 95 585/65 = 9
5. What goes in the first blank? In other words, what is the EV?
   a) 5 b) 6 c) 15 d) 50 e) 293 always expect half heads
6. What goes in the second blank? In other words, what is the SE?
   a) 2 b) 6 c) 15 d) 18 e) 30

SE/$\sqrt{n}$ = 6/$\sqrt{9}$ = 6/3 = 2

Questions 7-12 pertain to the following situations:

7. A gambler plays roulette 200 (= n) times betting a $1 on the numbers 7 and 11 each time. If the ball lands on 7 or 11 he wins $17, if it lands on any other number he loses $1. This corresponds to taking the sum of _____draws _____replacement from the corresponding box model?
   a) 200; with
   b) 200; without
   c) 38; with
   d) 17; with
   e) 17; without



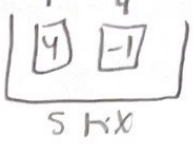
8. What is the appropriate box model for the scenario in Question 7? Remember, a roulette wheel has 38 slots.
   a) The box has 200 tickets, 2 marked "17" and 198 marked "-1".
   b) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.
   c) The box has 38 tickets, one marked "7", one marked "11" and the rest marked "0".
   d) The box has 38 tickets, 2 marked "17" and 36 marked "-1".
   e) The box has 38 tickets, 1 marked "35" and 37 marked "-1".

9. A multiple-choice test has 50 (= n) questions. Each question has 5 possible answers, only 1 of which is correct. Each correct answer is worth 4 points and 1 point is deducted for each incorrect answer. Suppose you guess at random on all 50 questions and your score is computed. This corresponds to taking the sum of _____draws_____ replacement from the corresponding box model?
   a) 50; with
   b) 50; without
   c) 4; with
   d) 4; without
   e) 125; with

10. What is the appropriate box model for the scenario in Question 9?

a) The box has 50 tickets, five tickets are marked "1" and 45 are marked "0"
b) The box has 5 tickets, one marked "1" and four marked "0"
c) The box has 5 tickets, one marked "4", and four marked "-1/4"
d) The box has 5 tickets, one marked "4", and four marked "-1"
e) The box has 50 tickets, one marked "4", and the rest marked "-1"

correct ↓ incorrect ← choices



11. You roll a die 30 times and count the number of 4s. This corresponds to taking the sum of _____draws _____replacement from the corresponding box model?

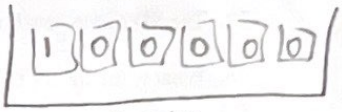
a) 30; with
b) 30; without
c) 100; with
d) 2; with
e) 2; without

n = 30

1-0 box: only care if we get a 4 or not

12. What is the appropriate box model for the scenario in Question 11?

a) The box has 6 tickets: 1 marked "1" and 5 marked "0".
b) The box has 6 tickets: 1 marked "4" and 5 marked "0"
c) The box has 6 tickets: one each of 1,2,3,4,5,6.
d) The box has 30 tickets: 5 each of 1,2,3,4,5,6.
e) The box has 30 tickets: one marked "4" and the rest marked "0"



Questions 13-20 pertain to the following situation: 100 draws are made at random with replacement from a box that has 4 tickets: 1 3 3 9

avg = $\frac{1+3+3+4}{4} = \frac{16}{4}$ = 4

13. What is the largest possible sum of the 100 draws?
a) 4 b) 100 c) 300 d) 400 **e)** 900

9 × 100 = 900

14. What is the smallest the sum can be?
a) 4 **b)** 100 c) 300 d) 400 e) 900

1 × 100 = 100

15. What is the EV for the sum of the draws?
a) 900 b) 600 **c)** 400 d) 300 e) 100

= n × avg
= 100 × 4 = 400

16. What is the SE for the sum of the draws? (The SD of the box is 3)
a) 0.5 b) 300 c) 100 **d)** 30 e) 900

= SD × $\sqrt{n}$ = 3 × $\sqrt{100}$
= 3 × 10 = 30

Continue to refer to the box above. For the next 3 questions, use the normal curve to estimate the *chance* that the sum of the draws is less than 445.

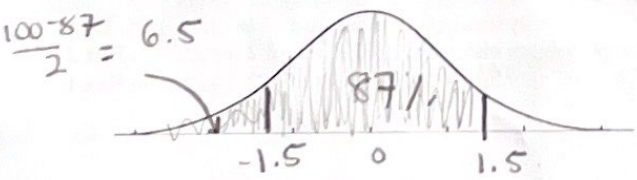
17. What is the z-score?
a) -1.5 b) 0.35 c) 0 d) 0.65 **e)** 1.5

$z = \frac{\text{val} - EV}{SE} = \frac{445 - 400}{30} = 1.5$

18. Mark your z-score on the normal curve! Do you shade to the left or to the right?
a) To the right **b)** To the left c) It doesn't matter

less than 445

19. What is the chance that the sum of the draws is less than 445?
a) 93.5% b) 87% c) 6.5% d) 66% e) 3.5%



$\frac{100-87}{2} = 6.5$

total shaded
= 87 + 6.5
= 93.5%

For the next two questions, think about drawing tickets out of the box from the previous page (shown here) and looking at the percent of 3s. [1] [3] [3] [9]

20. What is the expected value of the percent of 3's in 100 draws? = % in box

a) 4 b) 75 c) 50 d) 25 e) 100

To find the SE for the percent of 3's in 100 draws, you'll need to draw a new box!

21. The new box has….

[0] [1] [1] [0]

4 tix

both answers accepted due to a typo

a) 4 tickets: 1 marked "1" and 3 marked "0"
b) 4 tickets: 1 marked "0" and 3 marked "1"
c) 4 tickets: 2 marked "1" and 2 marked "0"
d) 2 tickets: 1 marked "1" and 1 marked "-1"
e) 4 tickets: 1 marked "1" and 3 marked "-1"

22. The SD of the new box is? a) 0 b) 3 c) 0.5 d) 1 e) 0.43 $= |1-0|\sqrt{.5 \times .5} = 0.5$

23. What is the SE for the percent of 1s? (typo) a) 4.3 b) 0.5 c) 0.05 d) 5 e) 10

SE% 1s: $SD = |1-0|\sqrt{.75 \times .25} = 0.43$

$SE\% = \frac{0.43}{\sqrt{100}} \times 100 = 4.3\%$

SE% 3s: $SE\% = \frac{0.5}{\sqrt{100}} \times 100 = 5\%$

Questions 24-28 pertain to the 2 boxes and 5 histograms below:

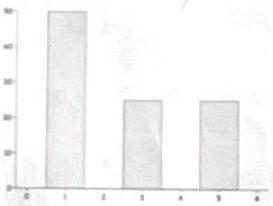
Box 1

[-1] [0] [1]

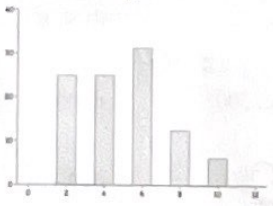
Box 2

[1] [1] [3] [5]

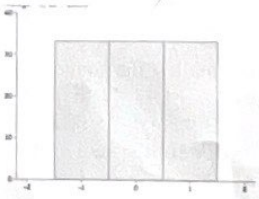
Histogram A



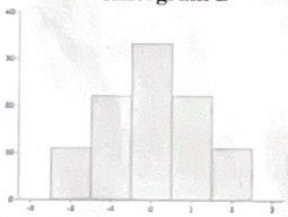
Histogram B



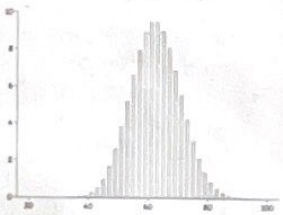
Histogram C



Histogram D



Histogram E



look @ x-axis: possible outcomes

Choose HISTOGRAM A, B, C, D, or E below to make the statements true.

24. Histogram a) A b) B c) C d) D e) E is the probability histogram for the **contents** of Box 2.
25. Histogram a) A b) B c) C d) D e) E is the probability histogram for the **contents** of Box 1.
26. Histogram a) A b) B c) C d) D e) E is the probability histogram for the **sum** of **25** draws from Box 1.
27. Histogram a) A b) B c) C d) D e) E is the probability histogram for the **sum** of **2** draws from Box 2.
28. Histogram a) A b) B c) C d) D e) E is the probability histogram for the **sum** of **2** draws from Box 1.

Questions 29 and 30 pertain to this situation: Suppose 50% of the households in the city of Chicago have school age children. You would expect 50% of the 400 households in the sample to have school age children with a SE for the sample % of 2.5%. Use this information and the normal curve to figure the chance that the percent of the sample households in Chicago that have school age children will be less than 49%. value

$= \frac{\text{val} - EV}{SE} = \frac{49-50}{2.5}$

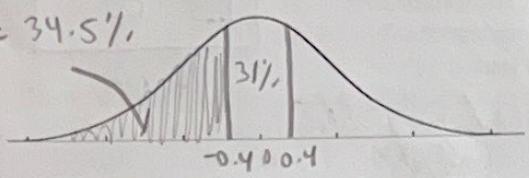
29. What's the z-score?

a) -0.5 b) 0.4 c) 0 d) -0.4 e) 0.5

30. What is the chance that the percent of the sample households in Chicago that have school age children will be less than 49%?

a) 65.5% b) 50% c) 34.5% d) 78% e) 31%

$\frac{100-31}{2} = 34.5\%$



shaded area

Questions 31-33 pertain to this situation: A political website conducts a public opinion poll daily called Quick Vote. Any Internet user can go to the website and cast their vote. On November 1st the Quick Vote question was: "Do you think the COVID-19 pandemic is over?" 8,900 people responded, 90% of people answered YES and the rest answered NO.

poll wasn't random

31. The main problem with this sample is:

a) Selection Bias since the people selected themselves b) Bias in the wording c) Sample Size

32. What is the SE for the percentage of YES's?

a) Impossible to calculate b) 0.35 c) 0.5 d) 50 e) 90

33. Does this poll accurately represent what all US adults think about this question? a) No b) Yes

Questions 34-35 pertain to the following scenario: A poll is taken in a city of population 100,000 (City A). A simple random sample of 1,000 is chosen and polled. Another poll is to be taken in the same way from another city (City B) with a population 100 times bigger (10 million people).

same sample size ↓ same accuracy

chocolate milk thm

34. In order to obtain the same accuracy as City A, the sample size in City B should be:

a) 100 b) 1,000 c) 100,000 d) 100 e) Not enough information to determine

35. If I wanted City B to have more accuracy than City A, the sample size should be:

a) Impossible to tell b) Decreased c) Increased d) Kept the same

↑ sample size = ↑ accuracy

Questions 36-38 pertain to this situation: A Fox News Poll asked a random sample of 900 adults nationwide the following question: "Do you personally believe in the existence of the Devil?" 71% of the people in the sample answered "YES".

z = 1.65

36. The SE of the % of people in the sample who said "YES" is about 1.5%. An approximate 89% confidence interval for the percentage of all American adults who believe in the Devil is:

a) 68.6%-73.4% b) 69.4%-72.4% c) 69.5%-72.5% d) Impossible to calculate

$EV \pm (z)(SE) = 71 \pm (1.6)(1.5) = 71 \pm 2.4 \rightarrow 68.6 - 73.4$

37. If the researcher increased the sample size to 8100 people, the length of an 89% confidence interval would…

a) Be divided by 3 b) Be multiplied by 9 c) Be multiplied by 3 d) Be divided by 9 e) Be multiplied by 81

law of avgs: from 900 people → 8100 people, n = 9

$SE_{new} = \frac{SE}{\sqrt{n}} = \frac{SE}{\sqrt{9}} = \frac{SE}{3}$ SE is divided by 3, and CI width does whatever SE does!

.92 .08 [1 0] .71 .29 [1 0]

38. In the same poll of 900 people, 92% answered "Yes" to the question: "Do you personally believe in the existence of God?" Would the SE of the % of people in the sample who said "YES" to this question still be 1.5%?
a) Yes, it would be exactly the same (b) No, it would be smaller c) No, it would be bigger

smaller SD → smaller SE

Questions 39-41 pertain to this situation: Suppose you are playing a game similar to roulette, except now the wheel has 50 slots instead of 38, each slot numbered 1-50. If you bet $1 on the number "3" and it comes up 3, you win $42; otherwise, you lose $1. What is the box model for your total winnings playing this gambling game 75 times, betting $1 each time?

39. Describe this box.
a) The box has 2 tickets: 1 marked "42" and 1 marked "-1"
b) The box has 2 tickets: 1 marked "1" and 1 marked "0"
c) The box has 75 tickets: 1 marked "42" and the rest marked "-1"
d) The box has 50 tickets: 3 marked "1" and 47 marked "0"
(e) The box has 50 tickets: 1 marked "42" and 49 marked "-1"

1 49 [42 -1] 50 tix

40. The average of this box is? $(42 \times 1) + (49 \times -1) / 50 = -7/50$
a) 42/50 b) 7/50 c) 0 (d) -7/50 e) -1/50

41. The SD of this box is?
a) 0.5 b) 0.14 (c) 6.02 d) 5.57 e) Impossible to calculate

$|42 - -1| \sqrt{\frac{49}{50} \times \frac{1}{50}} = 6.02$

Questions 42-52 pertain to the following situation: A recent survey asked a random sample of 1600 college students nationwide the following question: "How many hours have you spent watching TikTok in the past month?" The sample average was 20 hours and the SD was 16.

42. What most closely resembles the relevant box model?
a) It has 1600 tickets marked with "0"s and "1"s.
b) It has millions of tickets marked with "0"s and "1"s, but the exact percentage of each is unknown.
(c) It has millions of tickets. On each ticket is written a number indicating the hours spent watching TikTok. The exact average and SD are unknown but are estimated from the sample.
d) It has 1600 tickets. The average of the tickets is 20 and the SD is 16

43. How many draws are made from the relevant box model? sample size
(a) 1600 b) 400 c) 100 d) 200 e) 16

44. Do you draw with or without replacement? (a) Without b) With dont draw same person twice

45. What is the SE of the sample average?
a) 640 (b) 0.4 c) 40 d) Impossible to calculate since the data does not follow the normal curve.

$= \frac{SD}{\sqrt{n}} = \frac{16}{\sqrt{1600}} = \frac{16}{40} = 0.4$

46. Suppose 100 researchers each took a random sample of 1600 college students and each computed 95% confidence intervals, about how many of the confidence intervals would **miss** the average number of hours all college students spent watching TikTok in the past month?
a) All of them (b) 5 c) 50 d) 95 e) None of them since the data doesn't follow the normal curve

95 capture → 5 miss

$z = 1.65$

47. Calculate a 90% confidence interval for the average number of hours all college students spent watching TikTok in the past month.
(a) $20 \pm 1.65 * 0.4$ b) $20 \pm 1.65 * 16$ c) $20 \pm 1 * 0.4$ d) $90 \pm 1.65 * 0.4$

$EV \pm (z)(SE) \rightarrow 20 \pm (1.65)(0.4)$

EV ± (z)(SE)

The researchers computed 3 confidence intervals: a 68% CI, an 80% CI & a 95% CI from the same sample of 1600.

48. The shortest one is the _____CI. **a)** 68% **b)** 80% **c)** 95% **d)** Impossible to determine

z=1

49. The longest one is the _____CI. **a)** 68% **b)** 80% **c)** 95% **d)** Impossible to determine

z=2

50. How would you interpret the 95% CI for the average number of hours all college students spent watching TikTok in the past month? The interval is 20 ± 0.8 hours.

definition of CI

a) 95% of college students watched 20 ± 0.8 hours of TikTok in the past month.
b) 95% of the time college students watch TikTok, they will spend 20 ± 0.8 hours watching.
c) We are 95% sure that the true average number of hours college students spent watching TikTok this past month is in the interval 20 ± 0.8 hours.
d) We are 95% sure that college students watched 20 ± 0.8 hours of TikTok in the past month.

51. If the study asked the question: "Think about all the times you've done something that you later regretted. What percent of those times was alcohol involved?" the relevant box model would contain tickets with:
a) Only 1s and 0s **b)** Numbers ranging from about 0 to 100 **c)** Not enough information

percents

52. If the study described above asked the 1600 students whether or not they have listened to Taylor Swift's newest album (Midnights), the relevant box model would contain tickets with
a) Only 1s and 0s **b)** Numbers ranging from 0 to 100 **c)** Not enough information

yes/no

Questions 53-54 pertain to the following situation: Say that my husband, Steve, wanted to run for mayor of Champaign. For a pre-election poll in a close race, we may want a 95% confidence interval with a small margin of error.

z=2

53. Estimate how many people you'd need to poll to get a 95% confidence interval with only a 4% margin of error. **(Assume the SD of the population is around 0.5. Show work and circle answer)**
a) 95 **b)** 625 **c)** 4057 **d)** 9604 **e)** 10,000

$n = \left(\frac{100 \times SD \times z}{ME}\right)^2 = \left(\frac{(100)(2)(.5)}{4}\right)^2 = 625$

54. Estimate how many people you'd need to poll to get a 95% confidence interval with only a 1% margin of error. **(Assume the SD of the population is around 0.49. Show work and circle answer)**
a) 25 **b)** 98 **c)** 1407 **d)** 9604 **e)** 10,000

$n = \left(\frac{100 \times 2 \times 0.49}{1}\right)^2 = 9604$

Questions 55-58 to a 0-1 box.

55. The SD of a 0-1 box **CAN** be negative. **a)** False **b)** True

56. The largest that the SD of a 0-1 box can be is: **a)** 0 **b)** 0.2 **c)** 0.5 **d)** 1 **e)** 2

57. The smallest that the SD of a 0-1 box can be is: **a)** 0 **b)** 0.2 **c)** 0.5 **d)** 1 **e)** 2

58. The SD of a 0-1 box is largest when we have _____ % zeros & _____ % ones.
a) 0; 100 **b)** 100; 0 **c)** 25; 75 **d)** 75; 25 **e)** 50; 50

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