STATISTICS 100 EXAM 3 Fall, Key **Fall 2017**

PRINT NAME _____ _____ _____
(Last name) (First name) net ID (email)

CIRCLE SECTION: L1 (Flanagan 12:30pm) L2 (Flanagan 3:30pm) S1 (Danielle Sass) Online

(This page is worth 1 point. It is graded on writing your name and net id clearly and circling section.)

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.
SHOW WORK when requested, otherwise no credit.
Do NOT use scrap paper. **NO FORMULA SHEETS ALLOWED!**

Make sure you have all 7 pages including the normal table (11 problems).

For questions using the normal table, you may "round" z scores and percents to fit the closest line on the normal table and you may round percents on the table to the nearest whole number.

DO NOT WRITE BELOW THIS LINE

The numbers written in each blank below indicate how many points you missed on each page. The numbers printed to the right of each blank indicate how many points each page is worth.

Page 1 _____18

Page 2 _____16

Page 3 _____20

Page 4 _____14

Page 5 _____14

Page 6 _____17

Cover Page_____1

Total Score _____

There is NO CLASS tomorrow!

Scores will be posted on Compass by Friday morning and exams will be returned in class after Fall Break.

Online students may pick up their exam in 23 Illini Hall during office hours.

CONFLICT
Last names: L1/L2: A-R
141 Wohlers
(77 Exams)

WARNING-- Clear evidence of ANY cheating on this exam will result in a 0.

Page 1- FALL)

Statistics 100 Exam 3 FALL, 2017

A slacker student has 4 finals. Each final consists of 100 multiple-choice questions. He knows nothing so he decides to randomly guess on every question so he can complete each final in less than 5 minutes.

i) *(4 pts.)* To compute the expected value (EV) for the student's score for each final, you may need additional information. Which of the following do you need to know? **Circle** "Yes" if needed or "No" if not.

a)	How many points are awarded or deducted for each choice.	***Circle one:***	(Yes)	No
b)	How much time is allotted for the exam.	***Circle one:***	Yes	(No)
c)	How many students are taking each final.	***Circle one:***	Yes	(No)
d)	How many choices there are for each question.	***Circle one:***	(Yes)	No

ii) *(2 pts.)* Randomly guessing on all 100 questions corresponds to drawing 100 times with replacement from the appropriate box model. **(Fill in the first blank with a number and the second with either "with" or "without".)**

iii) *(4 pts.)* For **a-d** match the Final exams to their corresponding box models ***Use each box model exactly once.***

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

a) 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… (i) A ii) B iii) C iv) D

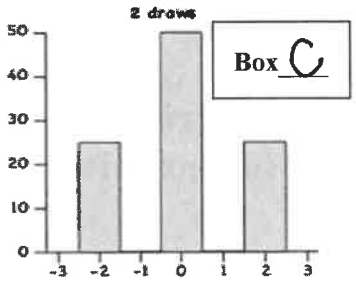
b) Final 2- Each question has 3 choices, one is the best answer and awarded 2 pts, one is a mediocre answer and awarded 1 pt. and one is a wrong answer and awarded no points. This corresponds to Box… i) A (ii) B iii) C iv) D

c) Final 3--Each question is a true/false question. Your score is the number of answers you get right. This corresponds to Box… i) A ii) B iii) C (iv) D

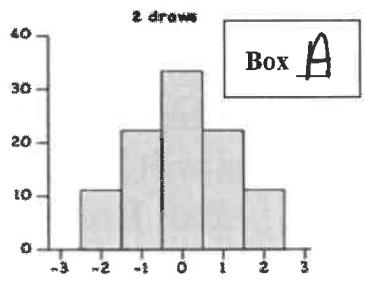
d) Final 4-Each question is a true/false question. Your score is the number of answers you get right minus the number of answers you get wrong. This corresponds to Box… i) A ii) B (iii) C iv) D

iv) *(4 pts.)* 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 (iii) above.** For each histogram identify the appropriate Box (A, B, C, or D). ***Use each box model exactly once.***

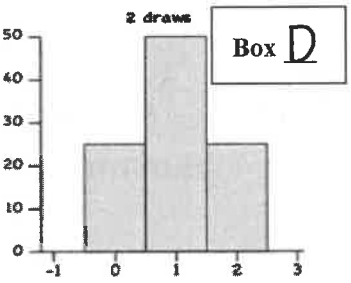
Histogram I



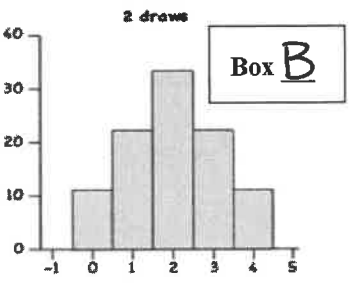
Histogram II



Histogram III



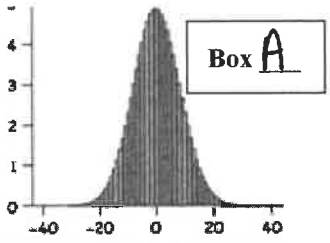
Histogram IV



v) *(4 pts.)* 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 part (iii) above.** For each histogram identify the appropriate Box (A, B, C, or D). ***Use each box model exactly once.***

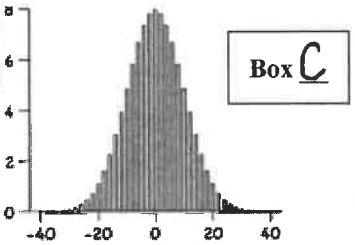
Histogram I

EV=0 SE=8.165



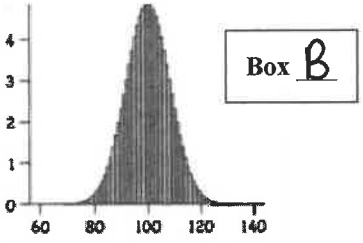
Histogram II

EV=0 SE=10



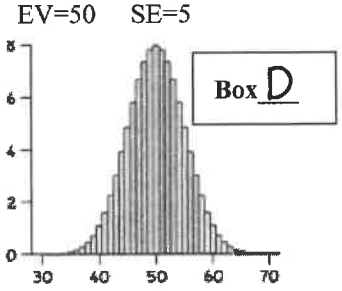
Histogram III

EV=100 SE=8.165



Histogram IV

EV=50 SE=5



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

Question 2 *(12 pts. total)*
25 draws are made at random with replacement from the box containing 4 tickets: [-4] [0] [0] [12]

a) *(2pts.)* The ***smallest*** the sum of the **25** draws could possibly be is **-100** and the ***largest*** is **300**.
(Fill in the 2 blanks above with the correct numbers.)

b) *(2pts.)* What is the **EV** (expected value) of the **sum** of the **25** draws? (Show work, circle answer.)

$EV_{sum} = n \times \text{avg of box} = 25 \times 2 = (50)$

c) *(2pts.)* What is the **SE** (Standard Error) of the **sum** of the **25** draws? **(SD of box = 6)** (Show work, circle answer.)

$SE_{sum} = \sqrt{n} \times \text{SD of box} = \sqrt{25} \times 6 = (30)$

d) *(1 pt.)* What is the **EV** of the **average** of the **25** draws? **2** (no work is necessary)

e) *(2 pts.)* What is the **SE** of the **average** of the **25** draws? **(SD of box = 6)** (Show work, circle answer.)

$SE_{avg} = \frac{\text{SD of box}}{\sqrt{n}} = 6/\sqrt{25} = (\frac{6}{5} \text{ or } 1.2)$

f) 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 0's you get. What are the **EV** and the **SE** of the **percent** of 0's in **25** draws? ***(Hint: draw a new box)***

i) *(1 pt.)* EV of the ***percent*** of 0's in **25** draws = **50** % do not accept 0.5

ii) Compute the SE% of 0's in 25 draws in two steps.

a) *(1 pt.)* First, what's the SD of the new box?
i) 0.2 ii) 0.433 iii) 0.5 iv) 1 v) 3

$SD = |1-0| \sqrt{2/4 \times 2/4} = 0.5$

b) *(1 pt.)* Now, use that to find the SE% of 0's in 25 draws. Show work below and circle your answer.

$SE_{\%} = \frac{SD}{\sqrt{n}} \times 100 = \frac{0.5}{\sqrt{25}} \times 100 = (10\%)$ continued error from a)

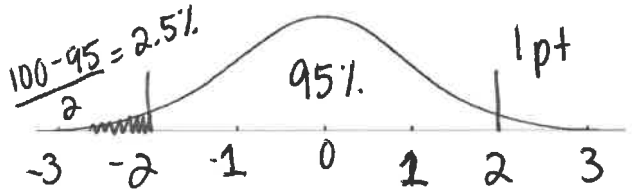
Question 3 *(4 pts. total)*

a) Use the normal approximation to find the **chance** that the sum of 400 draws will be **below 1880**?
The EVsum= 2000 and the SEsum= 60 for 400 draws.

i) *(2 pts.)* First calculate the Z score. ***Show work. Circle answer.***

$Z = \frac{1880-2000}{60} = (-2)$ -1 for forgetting negative

ii) *(2pts.)* Now mark the Z score accurately and ***shade the area that represents the chance of getting below 1880.***
Round the middle area given in the table to the nearest whole number. continued error from i)



Chance = **2.5** % 1 pt
accept 2.4-2.5%.

Question 4 *(8 pts. Total- 2 pts. each)*
Fill in the following chart for the EV and SE of the number (sum) and the percent of heads in 400 tosses of a fair coin, the first row is done for you for 4 tosses:

n = # of tosses	EV_{sum}	SE_{sum}	$EV_{\%}$	$SE_{\%}$
4 ×100	2	1 ×10	50%	25% ÷10
400	200 2pts	10 2pts	50% 2 pts	2.5% 2 pts

Question 5 *(12 pts. total)*
A gambler plays roulette **400** times betting \$1 on the 3 numbers: 17, 21, and 29 each time. If the ball lands on 17, 21, or 29 the gambler wins \$11, if the ball lands on any of the other 35 numbers the gambler loses \$1.
The roulette wheel has 38 slots numbered 1-36, 0 and 00.

a) *(2 pts.)* Which is the appropriate box model?
Circle one:

i) The box has 50 tickets: 21 marked "17" and 29 marked "-1"

ii) The box has 38 tickets: one each of 1, 2, 3, ..., 36, 0, and 00.

iii) The box has 38 tickets: 3 marked "11" and 35 marked "0"

iv) The box has 38 tickets: 1 marked "17",1 marked "21", 1 marked "29" and 35 marked "-1"

v) The box has 38 tickets: 3 marked "11" and 35 marked "-1"

b) *(1 pt.)* How many draws from the box? 400

c) *(1 pt.)* The draws are made .... ***Circle one***: **i)** without replacement **ii)** with replacement

d) *(2 pts.)* What is the average of the box? **Write your answer as a fraction**. (Show work, circle answer.)

$$\text{avg} = \frac{3(11)+35(-1)}{38} = \frac{-2}{38}$$ accept -0.05 or $-\frac{1}{19}$

e) *(2 pts.)* What is the SD of the box? ***Show work. (Hint: Use short-cut formula.) Circle answer. Round your answer to 2 decimal places.***

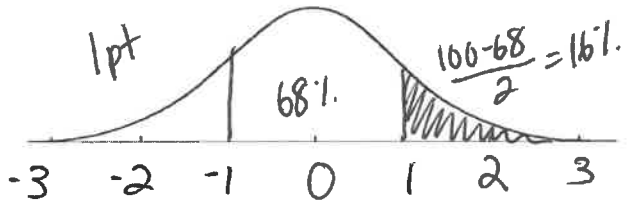
$$SD = |11-(-1)|\sqrt{\tfrac{3}{38}\times\tfrac{35}{38}} = 3.24$$ accept 3.23

f) Use the normal approximation and the fact that the **EV = \$ - 21** and the **SE = \$65** (approximately) to figure the chance that the gambler will win more than **\$44** in 400 plays.

i) *(2 pts.)* First calculate the Z score. ***Show work. Circle answer.***

$$Z = \frac{44-(-21)}{65} = 1$$

ii) *(2 pts.)*Now mark the z score on the curve, shade the area representing the chance of winning **more** than \$44 in 400 plays.
(Round the middle area given in the table to the nearest whole number.)



Chance = 16 **%**
1pt
accept between 15-16%.

continued error from i)

Question 6 pertains to the following situation: *(5 pts. total)*
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	48%	52%	1,001
WKRN.com Poll	60%	40%	2,550
Bonus Survey	18%	82%	912

a) *(2 pts.)* Which poll best reflects how all US adults would answer this question?
Choose one:
(i) The Marist random poll because the sample was ***randomly*** selected from the entire US adult population.
ii) The Stat 100 Bonus Survey results because we know it was anonymous.
iii) The WKRN Poll because it has the largest sample size.

b) *(3 pts.)* For each poll listed below, is it possible to calculate a 95% Confidence Interval for how all US adults would respond to the question?

i)	Marist Poll	***Circle either:***	(Yes)	or	No
ii)	Bonus Survey	***Circle either:***	Yes	or	(No)
iii)	WKRN.com Poll	***Circle either:***	Yes	or	(No)

Question 7 *(9 pts. total)*
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%

a) *(2 pts.)* The sample size for each question is the same. Is the SE of the sample percent the same for each of these questions?
Circle one: **i)** Yes
ii) No, the SE would be largest for belief in Black Cats changing one's luck and smallest for Haunted Houses.
(iii) No, the SE would be largest for belief in Haunted Houses and smallest for Black Cats changing one's luck.

b) *(2 pts.)* Which question has a bigger SE—belief in Ghosts or belief in Demons, or are they the same?
i) Ghosts **ii)** Demons (iii) They're exactly the same.

c) Fill in the blanks below to construct the correct Confidence Intervals (Hint—Use the normal table.)

i. *(1pt.)* An approximate **80%** Confidence Interval for the % of all US adults who would say they believe in **Ghosts** is

$$40\% +/- \underline{1.3} \times \frac{\sqrt{0.4 * 0.6}}{\sqrt{1111}} \times 100\%$$

ii. *(1pt.)* An approximate **95%** Confidence Interval for the % of all US adults who would say they believe in **Ghosts** is

$$40\% +/- \underline{2} \times \frac{\sqrt{0.4 * 0.6}}{\sqrt{1111}} \times 100\%$$

accept rounded & exact answers

iii. *(1pt.)* An approximate **99%** Confidence Interval for the % of all US adults who would say believe that people can become possessed by **Demons** is … (Be sure to use the closest line to 99% on the normal table.)

$$40\% +/- \underline{2.6} \times \frac{\sqrt{0.4 * 0.6}}{\sqrt{1111}} \times 100\ \%$$

d) *(2pts.)* If the researcher increased the sample size from 1111 to 9999 then the width of each confidence interval above would…
Circle one: **i)** be multiplied by 3 **ii)** be multiplied by 9 (iii) be divided by 3 **iv)** be divided by 9 **v)** stay the same

Question 8 *(10 pts. total)*
1100 students will take this exam. Suppose tomorrow at the grading meeting, I randomly pick 64 exams to grade to estimate the average of all 1100 exams. The sample of **64** exams has an **average = 87 with a SD = 12.**

a) *(2 pts.)* Which most closely resembles the relevant box model?
Circle one:

i) The box has 1100 tickets marked with "1"s and "0"s
ii) The box has 64 tickets marked with "1"s and "0"s
iii) The box has 64 tickets, marked with numbers ranging from about 20 to 100.
(iv) The box has 1100 tickets, marked with numbers ranging from about 20 to 100, the exact average is unknown but estimated from the sample.
v) The box has 1100 tickets with an average of 87 and a SD of 12.

b) *(1 pt.)* How many draws from the box? 64

c) *(1 pt.)* The draws are made .... ***Circle one***: **i)** with replacement **(ii)** without replacement

d) *(2 pts.)* The best estimate for the average of all 1100 exams is 87.

e) *(2 pts.)* What is the SE of the sample average? Show work. Circle answer.

$$SE_{avg} = \frac{SD}{\sqrt{n}} = \frac{12}{\sqrt{64}} = \left(\frac{12}{8} \text{ or } 1.5\right)$$

f) *(2 pts.)* 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?

Choose one:

i) No, if the data does not follow the normal curve, it's never possible to construct confidence intervals
ii) No, it's not possible to construct confidence intervals for averages.
(iii) 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.

Question 9 *(4 pts. total)*
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.

a) *(2 pts.)* Other things being equal, to achieve the same level of accuracy in the 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.

i) 25 times smaller than **ii)** 5 times smaller than **(iii)** the same as **iv)** 5 times larger than **v)** 25 times larger than.

b) *(1 pt.)* 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)

i) 400 **ii)** 625 **iii)** 1111 **(iv)** 2500 **v)** 10,000 $n = \left(\frac{200 \times 0.5}{2}\right)^2 = 2500$

c) *(1 pt.)* How many people would you have to poll in the Illinois to get a 95% Confidence Interval with a Margin of Error of **5%?** (Assume the SD of the population is close to 0.5)

(i) 400 **ii)** 625 **iii)** 1111 **iv)** 2500 **v)** 10,000 $n = \left(\frac{200 \times 0.5}{5}\right)^2 = 400$

Question 10 *(9 pts. total)*
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?"
1,637 of the respondents were White, 376 were Black, and 218 were Hispanic. Here are the results.

	Yes, have	No, have not	Sample Size
All US Adults	16%	84%	2,231
White Adults	10%	90%	1,637
Black Adults	35%	65%	376
Hispanic Adults	28%	72%	218

a) *(2 pts.)* True or False?
The reason a larger percentage of Whites answered "No" than Blacks (90% vs. 65%) is because the White sample size was larger than the Black sample size (1,637 vs. 376). ***Circle one:*** **i)** True **ii)** False

b) *(2 pts.)* What is the SE of the sample percent among Hispanics?
Choose one:
- **i.** It's not possible to calculate a SE for this sample because we don't know the size of the population.
- **ii.** It's not possible to calculate a SE for this sample because we don't know the SD of the sample.
- **iii.** The SE of the sample percent is about 2.5%
- **iv.** The SE of the sample percent is about 0.78%
- **v.** The SE of the sample percent is about 3%

$SE_{\%} = \frac{\sqrt{.28 \times .72}}{\sqrt{218}} \times 100 = 3\%$

c) *(2 pts.)* A 95% confidence interval for the percent of **all US adults** who would answer "yes, they have" is closest to:
i) (14%-16%) **ii)** (11%- 21%) **iii)** (10%-22%) **iv)** (14.5% - 17.5%)

$95\%\ CI = 16 \pm 2(0.776)$

$SE_{\%} = \frac{\sqrt{.84 \times .16}}{\sqrt{2231}} \times 100 = 0.776$

d) *(3pts.)* To which of the following populations can we also apply the above 95% confidence interval in (c)?
Circle Yes or No
- **i)** All Illinois adults — ***Circle either:*** Yes or (No)
- **ii)** All White adults in the US — ***Circle either:*** Yes or (No)
- **iii)** All Black adults in the US — ***Circle either:*** Yes or (No)

Question 11 *(8 pts. total)*
A nationwide Gallup poll asked a random sample of **1061** adults: "Are you afraid of public speaking in front of an audience?" **41%** of the people in the sample answered "YES".

a) *(2 pts.)* What most closely resembles the relevant box model?
Circle one:
- **i)** It has 1061 tickets, 41% are marked "1" and 59% are marked "0"
- **ii)** It has millions of tickets, exactly 41% are marked "1" and exactly 59% are marked "0"
- **iii)** It has millions of tickets marked "1" and "0", the exact percentage of each is unknown but are estimated from the sample to be 41% and 59% respectively.

b) *(2 pts.)* The poll reported a Margin of error= 3%. How did they get that number?
Circle one:
i) It's the SD of the sample **ii)** It's 2 x SE of the sample percent **iii)** It's the SE of the sample percent.

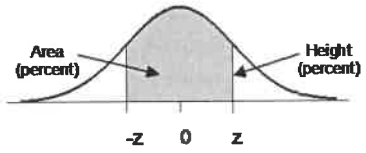
c) *(2 pts.)* We can be about 95% confident that if we polled all US adults the percent who would say they afraid of public speaking in front of an audience would be between 38 (41-3) % and 44 (41+3) %.

d) *(2 pts.)* Suppose 100 pollsters each randomly sampled 1,061 adults nationwide asking the same question. All 100 pollsters computed 90% confidence intervals to estimate the percentage of all US adults who would answer "Yes" to the question.

About how many of the 100 confidence intervals would miss the true population percentage? 10
accept 10%.

This is the end of the test. Go back and check your work.

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