

1 (14 pts)

first blank with the number of draws, the second with either "with" or "without" and the third with the letter corresponding to appropriate box model. Choose from the box models on the right. Use each box model exactly once.

- a) A die is rolled 5 times and the number of 4's is counted. This is like drawing 5 times w. th replacement from Box D
- b) A die is rolled 2 times and the sum of the spots is counted. This corresponds to drawing 2 times with replacement from Box C
- c) A true/false test has 6 questions. Suppose you randomly guess on all 6 questions and your score is the number of correct answers you get. This corresponds to drawing 6 times with replacement from Box B
- d) A true/false test has 10 questions. Suppose you randomly guess on all 10 questions and your score is the number of correct answers minus the number of incorrect answers. This corresponds to drawing 10 times with replacement from Box A

Box A -1 1

Box B 1 0

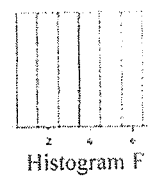
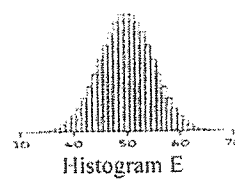
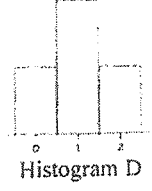
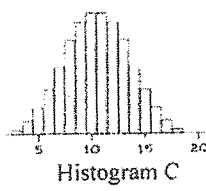
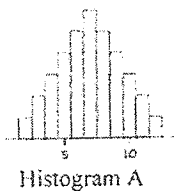
Box C 1 2 3 4 5 6

Box D 0 0 0 0 0 1

Don't forget to answer these 2 questions:

- e) Which box has the smallest SD? D (1 pt.)
- f) Which box has an SD=1 A (1 pt.)

Question 2 pertains to the 6 probability histograms below, which display chances for coin tosses and dice rolls. (6 pts.)  
Match the histogram with its description. Use each histogram exactly once.



- a) The probability histogram for counting the number of heads when tossing a fair coin... (3 pts.)  
i) twice is Histogram D ii) 3 times is Histogram B iii) 100 times is Histogram E
- b) The probability histogram for counting the total number of spots when rolling a fair die... (3 pts.)  
i) once is Histogram F ii) twice is Histogram A iii) 3 times is Histogram C

Question 3 (12 pts.)

A gambler plays roulette 400 times betting \$1 on the 2 numbers: 7 and 8. If the ball lands on either 7 or 8 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.

- a) Draw the appropriate box model in the box below. Label the 2 rectangular tickets inside the box with the correct numbers and write how many of each type of ticket in the blank below each ticket. (4 pts.)

\$17	\$-1
<u>2</u>	<u>36</u>

- i) What is the average of the box? Show work. Leave answer as a fraction. (2 pts.)

$$\frac{2(17) + 36(-1)}{38} = \frac{34 - 36}{38} = \frac{-2}{38}$$

- ii) What is the SD of the box? Hint: Use short-cut. Show work. Round answer to 2 decimals. (2pts.)

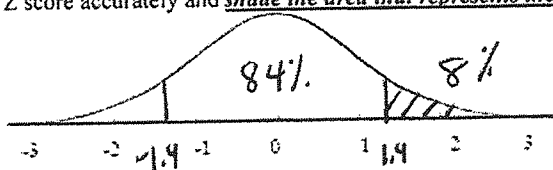
$$\sqrt{17^2 - (-1)^2} \sqrt{\frac{2}{38} \cdot \frac{36}{38}} = 18(.223) = 4.02$$

- b) Use the normal approximation and fact that the EV is about \$-21 and the SE is about \$80 to figure the chance that the gambler will win more than \$91 in 400 plays?

- i. First calculate the Z score. (Show work, circle answer.) (2 pts.)

$$Z = \frac{91 - (-21)}{80} = \frac{112}{80} = 1.4$$

- ii. Now mark the Z score accurately and shade the area that represents the chance of winning more than \$91. (2 pt.)



Chance = 8%  
(Round to nearest whole number.)

Question 4 (20 pts.)

25 draws are made at random with replacement from the box containing 6 tickets:  $\boxed{-1} \boxed{-1} \boxed{5} \boxed{5} \boxed{5} \boxed{17}$   
 (The AVERAGE of Box = 5 and the SD of the box = 6)

- a) The *smallest* the sum of the 25 draws could possibly be is -25 and the *largest* is 425 (2pts.)  
 (Fill in the 2 blanks above with the correct numbers.)

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

$$EV_{sum} = n \cdot ave = 25 \cdot 5 = \boxed{125}$$

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

$$SE_{sum} = SD \cdot \sqrt{n} = 6 \sqrt{25} = 6 \cdot 5 = \boxed{30}$$

- d) Use the normal approximation and your answers from (b) and (c) above to find the chance that the sum will be below 185?

- i) First calculate the Z score. Show work. Circle answer. (2 pts.) *c.e. from (b) and (c)*

$$Z = \frac{185 - 125}{30} = \boxed{2}$$

- ii) Now mark the Z score accurately and shade the area that represents the chance of getting below 185. Round the middle area given in the table to the nearest whole number.



$$\text{Chance} = \boxed{97.5\%} \quad (2pts.)$$

*-1. for 2.5%*

- e) What is the EV of the average of the 25 draws? 5 (no work is necessary) (2 pt.)

- f) What is the SE of the average of the 25 draws? 1.2 (SD of box = 6) (Show work, circle answer.) (2 pts.)

$$SE_{ave} = SD / \sqrt{n} = 6 / \sqrt{25} = 6 / 5 = 1.2$$

- g) Now suppose you draw at random with replacement from the same box above, but this time you're only interested in the percent of 5's you get. What is the EV and the SE of the *percent* of 5's in 25 draws? (Hint: draw a new box)

- i) EV of the *percent* of 5's in 25 draws = 50 % (1 pt.)

- ii) SE of the *percent* of 5's in 25 draws = 10 % Round your answer to 2 decimal places (2pts)  
 (Show work by computing the SD of the new box, then use it to calculate the SE %)

$$\boxed{1001110}$$

$$SD = \sqrt{3/6 \cdot 3/6} = \boxed{.5}$$

*1 pt for .5*

$$SE_{\%} = \frac{SD}{\sqrt{n}} \cdot 100\% = \frac{.5}{\sqrt{25}} \cdot 100\%$$

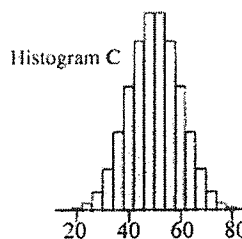
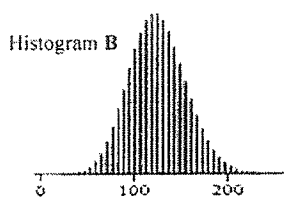
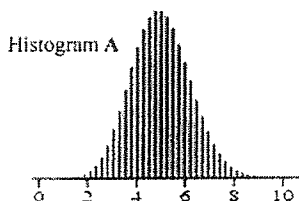
$$= \frac{.5}{5} \cdot 100\% = 10\%$$

*1 pt if all correct except .5*

- h) The 3 histograms below (in scrambled order) represent 25 random draws with replacement from the box at the top of the page.

**\*\*Fill in the 3 blanks below with either A, B, or C to correctly identify each histogram\*\*** (3 pts)

Histogram B is for the sum of the draws, A is for the average of the draws, and C is for the % of 5's drawn.



## Question 5 (6 pts.)

A Fox News Poll asked a random sample of 500 men and 500 women nationwide the following question: "Trying to be as honest as you can, are you more attracted to people by their bodies or their brains?"

24% of the women and 44% of the men in the samples answered "Bodies".

- a) An approximate 68% confidence interval for the percentage of all American women who would say they are more attracted to bodies than brains is: (2 pts.)

Choose one:

i)  $44\% \pm \frac{\sqrt{0.44 * 0.56}}{\sqrt{500}} \times 100\%$

ii)  $24\% \pm \frac{\sqrt{0.24 * 0.76}}{\sqrt{500}} \times 100\%$

iii)  $68\% \pm \frac{\sqrt{0.44 * 0.56}}{\sqrt{500}} \times 100\%$

iv)  $44\% \pm \frac{\sqrt{0.24 * 0.44}}{\sqrt{1000}} \times 100\%$

v)  $24\% \pm \frac{\sqrt{0.24 * 0.44}}{\sqrt{1000}} \times 100\%$

- b) The sample sizes for the men and women are exactly the same. Does that mean that the SE of the % of people in each sample who answered "Bodies" to the question is also exactly the same? (2pts)

Choose one:

i) Yes

ii) No, the SE in the men's sample is bigger.

iii) No, the SE in the women's sample is bigger.

iv) Not enough information is given to compare the SE's of the two samples

- c) If you combine the 2 samples into one sample of 1000 you get 34% answering "Bodies". Would a 68% confidence interval for the percentage of all American adults who say they're more attracted to bodies than brains be the same size as a 68% confidence interval computed for just men? In other words, would it be the same size as the confidence interval you computed in part (a)? (2 pts.)

Choose one:

i) No, the confidence interval for all Americans would be wider since the sample size is bigger.

ii) No, the confidence interval for all Americans would be narrower since the sample size is bigger.

iii) The confidence interval for all Americans would be exactly the same size since the sample is bigger but the SD is smaller, so the 2 effects would cancel out.

iv) Not enough information to determine.

## Question 6 (10 pts.)

A health survey asked a random sample of 1600 college students nationwide the following question: "How many alcoholic drinks have you consumed in the past 2 weeks?" The sample average was 18 drinks and the SD was 12.

- a) What most closely resembles the relevant box model? Circle one. (2 pts)

i) It has 1600 tickets marked with "0"s and "1"s.

ii) It has millions of tickets marked with "0"s and "1"s, but the exact percentage of each is unknown.

iii) It has millions of tickets. On each ticket is written a number indicating the number of drinks. The exact average and SD are unknown but are estimated from the sample.

iv) It has 1600 tickets. The average of the tickets is 18 and the SD is 12

- b) The draws are made \_\_\_\_\_ replacement. i) With ii) Without (2 pt.)

- c) What is the SE of the sample average? (2 pts)

Choose one:

i) 640

ii) 0.3

iii) 0.4

iv) 480

iv) Impossible to calculate since the data does not follow the normal curve.

- d) Suppose 100 researchers each took a random sample of 1600 college students and each computed 90% confidence intervals, about how many of the confidence intervals would cover the average number of drinks all college students consumed in the past 2 weeks?

Choose one:

i) All of them

ii) 95

iii) 90

iv) 5

v) 10

vi) None of them since the data doesn't follow the normal curve. (2 pts)

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

i) Which is the narrowest confidence interval?

Circle one:

68% CI

80% CI

95% CI

(1 pt)

ii) Which is the widest confidence interval?

Circle one:

68% CI

80% CI

95% CI

(1 pt)

## STATISTICS 100 EXAM 3

Page 4

Spring 2014

7 ↔ 9  
same 8

Question 7 (9) pertains to the following situation: (4 pts.) (9 on CAP)  
During the same week in January 2014, 2 polls asked the same question: "Do you favor or oppose raising the federal minimum wage to \$10 an hour?" The CBS News poll asked that question of a randomly selected sample of 1,644 adults nationwide, and the Gretawire Poll simply posted the question on its website <http://gretawire.foxnewsinsider.com> and allowed anyone who visited the website to cast a vote. Here are the results:

	Favor	Oppose	Unsure	Sample Size
CBS News Poll	72%	26%	2%	1,644
Gretawire Poll	18%	82%	0%	2,606

a) Which poll best reflects how all US adults would answer this question? (2pts.)

Choose one:

- i) The Gretawire Poll because it has the larger sample size.  
 ii) ☒ The CBS News random poll because the sample was randomly selected from the entire US adult population.  
 iii) The Gretawire Poll because it has a smaller SD.

b) For which poll is it possible to calculate a 95% Confidence Interval for how all US adults would respond to the question? (Assume if we collapsed the 3 response categories into 2 so it's possible to compute a  $SE_{\hat{p}}$  for those who would answer "Favor.") (2 pts.)

Choose one:

- i) Only the Gretawire Poll    ii) ☒ Only the CBS random Poll    iii) Both Polls    iv) Neither Poll

## Question 8 (14 pts.)

Suppose I wanted to determine the percentage of students in the UI Class of 2013 who are currently employed. There are approximately 7000 students in the UI Class of 2013. I randomly chose 100 students to survey, and 70% of them said they were employed.

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

- i) It has 7000 tickets, 70% marked "1" and 30% marked "0".  
 ii) It has 100 tickets, 70% marked "1" and 30% marked "0".  
 iii) It has 7000 tickets, with an average of 0.70, but the SD is unknown.  
 iv) ☒ It has 7000 tickets marked with "0"s and "1"s. The exact percentages are unknown but are estimated from the sample.

b) What is the expected value for the percent of all UI Class of 2013 students who are currently employed? 70% (2 pt)

c) What is the expected value for the percent of all men in the UI Class of 2013 who are currently employed? (2 pt)

- i) 70%    ii) 30%    iii) ☒ not enough information to determine

d) What is the SE for the percent of all UI Class of 2013 students who are currently employed? Round answer to 2 decimal places.  
 (Show work, circle answer.) (2 pts)

$$SE_{\hat{p}} = SD / \sqrt{n} * 100\% = \frac{\sqrt{.7 * .3}}{\sqrt{100}} * 100\% = \textcircled{4.58\%}$$

e) An approximate 80% confidence interval for the percentage of all UI Class of 2013 students who are currently employed is Choose one: (2 pts)

- i) 70% +/- 2 \* ( $SE_{\hat{p}}$ )    ii) ☒ 70% +/- 1.3 \* ( $SE_{\hat{p}}$ )    iii) 70% +/- 0.8 \* ( $SE_{\hat{p}}$ )    iv) 70% +/- 1 \* ( $SE_{\hat{p}}$ )

f) To double the accuracy of the survey, the sample size would need to ..... (doubling accuracy is the same as halving the  $SE_{\hat{p}}$ )

Choose one: (2 pts)

- i) be multiplied by 2    ii) ☒ be multiplied by 4    iii) be divided by 2    iv) be divided by 4    v) not possible to calculate

g) Suppose I was interested in the employment status of the approximately 70,000 UI students who graduated in the past 10 years. I'd have to randomly draw a new sample from all UI grads of the past 10 years. To keep the same SE as in part (d), should I keep the sample size about the same as before (100 people)?

Choose one: (2 pts)

- i) No, you should significantly increase it    ii) No, you should significantly decrease it    iii) ☒ Yes, keep it about the same

# STATISTICS 100 EXAM 3

Spring 2014

9 ← 7  
same 10

## Question 9 (7) (8 pts.)

(Question 7 on GAPS)

A CNN poll conducted in Feb 2014 asked a random sample of 1,010 US adults nationwide the following question: "Do you agree or disagree with the following statement: 'The government should work to substantially reduce the income gap between the rich and the poor'." 66% of the sample agreed with the statement.

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

- i) It has 1,010 tickets 66% marked "1" and 34 marked "0".
- ii) It has millions of tickets marked with "1"s and "0"s, but the exact percentage of each is unknown but are estimated to be 66% "1"s and 34% "0"s from the sample
- iii) It has millions of tickets marked, 66% marked "1" and 34 marked "0".
- iv) It has millions of tickets, on each ticket is an income ranging from 0 to billions of dollars.

b) The CNN poll above reported a Margin of error = 3% for a 95% Confidence Interval. How did they get that number? (2 pts.)

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) We can be about 95% confident that if we polled all US adults the percent who agree that "the government should work to substantially reduce the income gap between the rich and the poor" would be between 63 % and 69 %.

(2 pts.)

d) To which of the following populations can we also apply the above 95% confidence interval? (In other words which of the following could we substitute for "all US adults" in the above statement and be correct?). (2 pts.)

- i) All US Democrats
- ii) All Illinois adults
- iii) All US men
- iv) All US Republicans
- v) None of the

## Question 10 (6 pts.)

Suppose a survey organization is planning to take a random poll in Pennsylvania (population about 10 million adults) and a random poll in US (population about 250 million adults) to estimate the percent of adults at both the state and the national level who would support legislation to make Presidential election days federal holidays.

a) Other things being equal, to achieve the same level of accuracy in both polls, the number of people you'd have to poll in Pennsylvania is about \_\_\_\_\_ the number of people you'd have to poll in the whole US. (2 pts.)

- 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) About how many people would you have to poll in the US to get a 95% Confidence Interval with a Margin of Error of 1.0%? (Assume the SD of the population is close to 0.5) (1 pt)

i) 100

ii) 400

iii) 625

iv) 1111

v) 2500

vi) 10,000

$$\left(\frac{100}{10}\right)^2$$

c) About how many people would you have to poll in the Pennsylvania to get a 95% Confidence Interval with a Margin of Error of 3%? (Assume the SD of the population is close to 0.5) (1 pt)

i) 100

ii) 400

iii) 625

iv) 1111

v) 2500

vi) 10,000

$$\left(\frac{100}{3}\right)^2$$