STAT 110 **Name:** _____
Practice Exam 1
Fall 2017

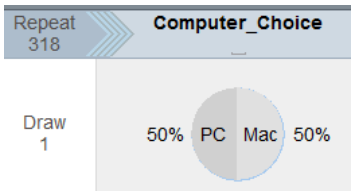
True/False Questions: Answer the following true/false questions.

Questions 1 - 8 concern the following scenario. A polling company surveyed a random sample of 439 adults nationwide who are employed either full-time or part-time. They were asked the following question: "*Are you completely satisfied with the amount of vacation time you receive*?" Of those surveyed, 53% answered "yes," and the poll reported a margin of error of ± 5% (this is the margin of error associated with a 95% confidence interval).

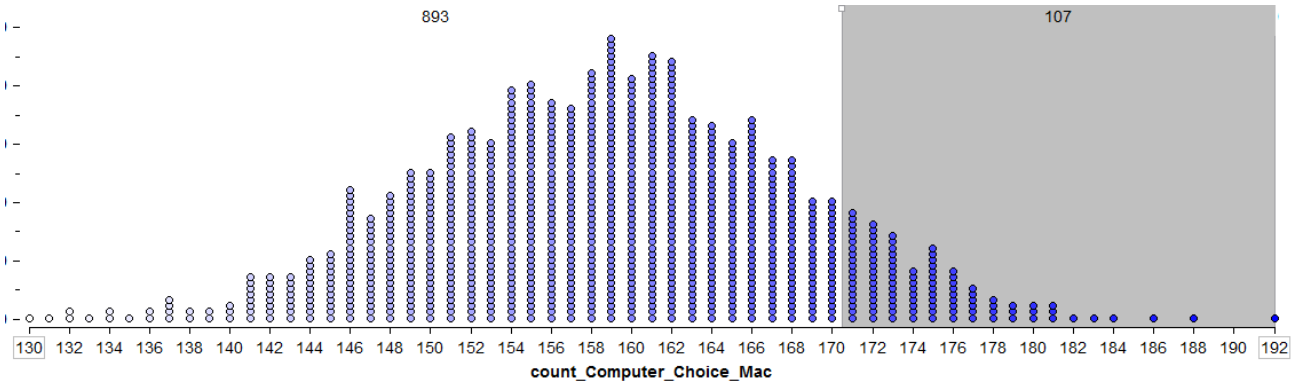
1. The data that this survey question yields is numerical because we can calculate the proportion of adults in the study that respond with "*yes*" (53%).	TRUE (FALSE)
2. The population of interest for this poll is *all* adults nationwide who are employed either full-time or part-time.	(TRUE) FALSE
3. The statement "Of those surveyed, 53% said they were completely satisfied" is an example of the use of *inferential* statistics.	TRUE (FALSE)
4. If researchers wanted to use these data to provide evidence that the majority (i.e., more than 50%) of employed adults nationwide are completely satisfied with the amount of vacation time they receive, their alternative hypothesis would be given as follows. H_a: The proportion of all employed adults nationwide who are completely satisfied with the amount of vacation time they receive is greater than 53%.	TRUE (FALSE)
5. Given the margin of error for this study, we can say that we are 95% certain the true proportion of all employed adults nationwide who are completely satisfied with the amount of vacation time they receive lies between 48% and 58% (i.e., 53% ± 5%).	(TRUE) FALSE
6. This survey result gives us evidence that more than half of *all* employed adults nationwide are completely satisfied with the amount of vacation time they receive.	TRUE (FALSE)
7. Suppose that these same survey results were used to construct a 90% confidence interval, instead. This interval would be narrower than the 95% confidence interval (that is, the margin of error would be smaller).	(TRUE) FALSE
8. Holding all else constant, if the polling company had increased the sample size by surveying 800 adults, instead, the margin of error would be larger than it was for this poll.	TRUE (FALSE)

Questions 9 - 13 concern the following scenario. A total of 318 students were asked, "Do you prefer a PC or a Mac?" Of these 318 students, 171 selected the PC ($\hat{\pi}$ = 171/318 = 53.8%). A supervisor in Tech Support asks the following question: "Is this enough evidence to say that the majority of WSU students, in general, prefer a PC over a Mac?" You decide to take a statistical approach to answer this question.

Consider the following outcomes from 1,000 simulations carried out in Tinkerplots. The spinner for this simulation was set up follows:



The number of students that chose the PC was recorded for each trial, and the results are shown below.



9. Using the 5% rule and the observed result from our study (i.e., 171 chose the PC), there is enough statistical evidence to conclude that the majority of all WSU students prefer a PC.	TRUE (FALSE)
10. If the spinner is not changed and many more simulations are carried out, the dots will become more and more centered around 171 (i.e., the observed result from our study).	TRUE (FALSE)
11. Having more dots at or above 171 would mean we would have *less* statistical evidence that students prefer a PC.	(TRUE) FALSE
12. The p-value for this investigation would be the proportion of dots at or above 171 on this graph.	(TRUE) FALSE
13. The binomial distribution with n = 318 and π = 0.50 would have approximately the same shape as the above plot.	(TRUE) FALSE

Multiple Choice Questions:

14. Suppose that researchers want to show that the majority (i.e., more than 50%) of registered nurses have been bullied in the workplace. They obtain a random sample of 50,000 registered nurses nationwide, and 25,300/50,000 = 50.6% report having been bullied in the workplace. The p-value for testing the research hypothesis is found to be 0.0037. These survey results are

a. Both statistically and practically significant.
b. Neither statistically nor practically significant.
c. Statistically significant but not practically significant.
d. Statistically non-significant but practically significant.

15. Suppose a recent poll was conducted to estimate the proportion of Minnesotans age 18-30 who feel that they are not represented by either the Republican or Democratic parties. The margin of error associated with this poll of 1,000 Minnesotans age 18-30 was about ± 3%. Now, suppose that you want to estimate the proportion of adults age 18-30 *nationwide* and that you want to achieve the same margin of error (± 3%) for your poll, as well. How many adults nationwide (age 18-30) would you need to randomly sample to achieve this margin of error?

a. Much more than 1,000
b. About 1,000
c. Much fewer than 1,000
d. It is impossible to tell without knowing the exact number of adults age 18-30 living in Minnesota and the exact number of adults age 18-30 living in the U.S., overall.

Questions 16 - 18 concern the following scenario. Recently, a company laid off 50 employees. You are investigating a claim that this company was discriminating against older persons (age 60 or older) in the layoff process (specifically, the concern was that too many older persons were laid off). Overall, 10% of the employees who were originally at risk for layoff were age 60 or older.

16. Let π = the probability that a person who was laid off was age 60 or older. Which are the appropriate null and alternative hypotheses for investigating the discrimination claim?

a. H_o: $\pi = 50\%$
   H_a: $\pi < 50\%$
b. H_o: $\pi = 50\%$
   H_a: $\pi > 50\%$
c. H_o: $\pi = 10\%$
   H_a: $\pi < 10\%$
d. H_o: $\pi = 10\%$
   H_a: $\pi > 10\%$

17. Suppose the result does turn out to be statistically significant (p-value = 0.0022). Which of the following is a valid conclusion?

a. An outcome at least as extreme as the observed result is <u>likely</u> to occur purely by chance assuming the company is <u>not guilty</u> of age discrimination.
b. An outcome at least as extreme as the observed result is <u>unlikely</u> to occur purely by chance assuming the company is <u>not guilty</u> of age discrimination.
c. An outcome at least as extreme as the observed result is <u>unlikely</u> to occur purely by chance assuming the company is <u>guilty</u> of age discrimination.

18. Now, suppose that of the 50 employees laid off, 33 were age 60 or older. The p-value associated with this observed result is 0.2369. Which of the following conclusions would be most correct?

a. The company is not discriminating against older persons in the layoff process.
b. The company is discriminating against older persons in the layoff process.
(c.) The company might be guilty of discriminating against older persons in the layoff process, but the sample size is not large enough to detect the discrimination effect in this particular case.

Questions 19 - 21 concern the following scenario. A friend claims that they can taste the difference between Diet Coke and Caffeine-free Diet Coke. You set up an experiment in which your friend blindly tastes a sample of each and identifies which is the Caffeine-free Diet Coke. The order in which they taste the samples is randomized, and this process is repeated a total of three times.

Trial 1	Trial 2	Trial 3
		

19. Suppose your friend correctly identifies the Caffeine-free Diet Coke on all three trials. You plan to use the binomial distribution to calculate the probability they would get all three correct *just by guessing*. What value would you use for π = the probability of correctly identifying the Caffeine-free Diet Coke?

a. $\pi = 1/4$
b. $\pi = 4/4$
(c.) $\pi = 1/2$
d. $\pi = 4$

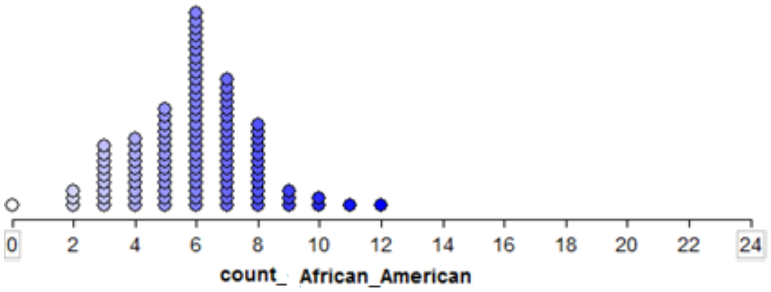
20. Recall that your friend correctly identifies the Caffeine-free Diet Coke on all three trials. You use the binomial distribution and find that the p-value is 0.13. Which of the following is the most correctly written conclusion?

a. There is convincing statistical evidence that your friend can taste the difference because they correctly identified the Caffeine-free Diet Coke in all three trials (100% of the time).
(b.) There is <u>not</u> enough convincing statistical evidence that your friend can taste the difference because this probability is above 0.05.
c. There is convincing statistical evidence that your friend can taste the difference because this probability is above 0.05.

21. Recall that the p-value for this experiment is 0.13. Which of the following statements is the most correct interpretation of this p-value?

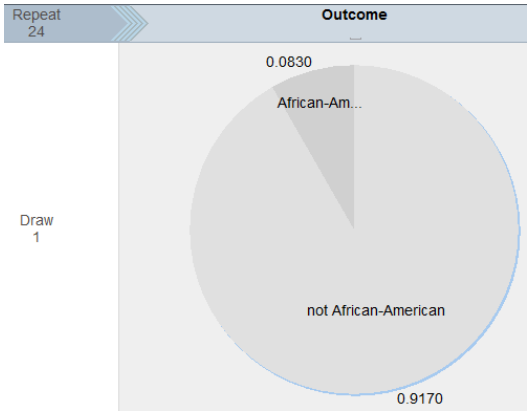
a. There is a 13% chance that your friend is guessing on each trial.
b. There is a 13% chance that your friend can taste the difference between Diet Coke and Caffeine-free Diet Coke.
(c.) There is a 13% chance your friend would get all three correct even if they are just guessing.
d. There is a 13% chance your friend would get all three correct, assuming they really can taste the difference between Diet Coke and Caffeine-free Diet Coke.

Questions 22 - 23 concern the following scenario. Suppose that 25% of a county's population consists of African-Americans. A pool of 24 potential jurors is selected for jury duty, and only 2 of the 24 are African-American ($\hat{\pi}$ = 8.3%). The county claims that the selection was truly random, but the defense attorney questions whether the selection process was biased against African-Americans. She carries out a simulation study with 100 trials in Tinkerplots to get an idea of what to expect if the selection process is fair. The number of African-Americans was recorded for each trial of the simulation, and the results are shown below.

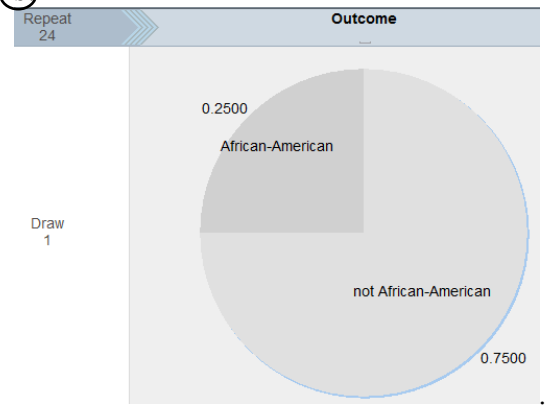


22. Which of the following spinners was used for this simulation study?

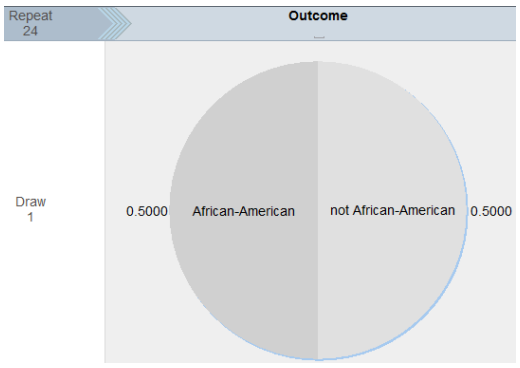
a.



(b.)



c.



23. Recall that of the 24 potential jurors selected, 2 were African American. Which of the following is the estimate of the p-value based on this simulation study?

(a.) The p-value would be about 4% because 4/100 dots are at 2 or below.
b. The p-value would be about 3% because 3/100 dots are right at 2.
c. The p-value would be about 99% because 99/100 dots are at 2 or above.
d. The p-value would be about 8.3% because 2/24 of the potential jurors were African-American.

Short Answer Questions:

Questions 24 - 27 refer to the following scenario. A study was conducted to investigate the preferences of female *Rhinogobius* fish. Each fish tank used in the study was set up so that it contained a small nest on one side and a large nest on the other side. A female was placed in the tank, and researchers recorded which nest she chose. Of the 17 females tested, 11 chose the large nest.

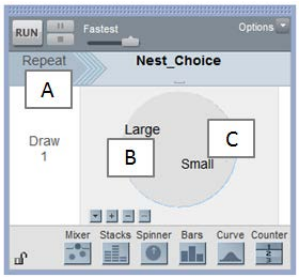
Research Question: Are the female *Rhinogobius* fish more likely to choose to lay their eggs in a larger rather than a smaller nest?

24. You have been asked to conduct a simulation study to determine what outcomes are likely to occur assuming that the fish have no real preference regarding nest size. Below, identify the appropriate values for a spinner in Tinkerplots that would allow you to do this.

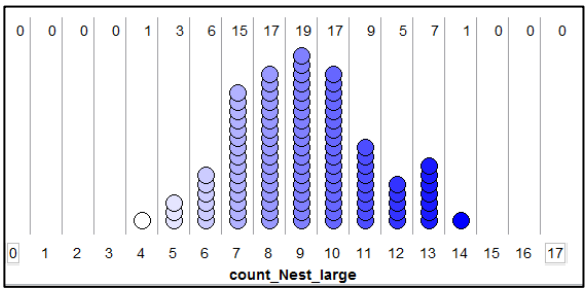
a. Box A contains the Repeat value. This should be set to __**17**____.

b. Box B sets the value for the proportion of fish that choose the large nest. This value should be set to _**50%**_.

c. Box C sets the value for the proportion of fish that choose the small nest. This value should be set to _**50%**_.



25. Suppose the results of 100 simulated trials are as shown in the following dotplot. Note that we kept track of the number that chose the large nest in each simulated trial.



Use this information to fill in the blank in the following statement: "In exactly __**17**__ of our 100 simulated trials, we observed 10 fish choose the large nest."

26. Recall that in the actual study, 11 fish chose the large nest. Use the dotplot from the previous question to estimate the p-value for this study.

The estimated p-value is __**22/100 = 0.22**__.

27. Which of the following is the most correct statement regarding our research question?

a. An outcome such as 11 or more fish choosing the larger nest happened more than 5% of the time in our simulation study; thus, there is not enough statistical evidence that female *Rhinogobius* fish are more likely to choose to lay their eggs in a larger rather than a smaller nest.

b. In the study, 11/17 = 65% of the fish chose the larger nest. We have statistical evidence that female *Rhinogobius* fish are more likely to choose to lay their eggs in a larger rather than a smaller nest simply because more than half of the fish in our study did so.

c. In the study, 11/17 = 65% of the fish chose the larger nest; since this value is greater than .05 (or 5%), we do not have statistical evidence that female *Rhinogobius* fish are more likely to choose to lay their eggs in a larger rather than a smaller nest.

Questions 28 - 30 concern the following scenario. In February of 2017, a Quinnipiac University poll surveyed a random sample of 1,155 registered voters nationwide. They were asked the following question, "*Do you think Senate Republicans were right or wrong to prevent for 10 months a vote on Merrick Garland, who was President Obama's nominee to the Supreme Court?"* Of those surveyed, 647 ($\hat{\pi}$ = 647/1155 = 56%) answered "*wrong*." The poll reported a margin of error of ± 2.9% (this is the margin of error associated with a 95% confidence interval).

28. Use Wald's method to verify the margin of error reported here (± 2.9%). You must show all of the work involved in this computation.

$$\mathbf{1.96}\sqrt{\frac{\mathbf{0.56(1-0.56)}}{\mathbf{1155}}}$$

29. Use the margin of error to report a 95% confidence interval for π = the proportion of *all* registered voters that felt the Senate Republicans were wrong. Give both the lower and upper endpoints of the interval.

0.56 – 0.029 = 0.531
0.56 + 0.029 = 0.589

30. Based on this confidence interval, is it fair to say that this poll provides evidence that more than half of all registered voters felt the Senate Republicans were wrong to prevent a vote on Garland? Explain why or why not.

Yes, since the entire CI sits above 50%, we have statistical evidence that more than 50% of all registered voters felt they were wrong.

31. Suppose that a commonly used pain medication is known to cause nausea in about 25% of the patients that use it. A pharmaceutical company has created a new drug with the same pain-killing properties but claims that patients who take it are less likely to experience nausea than with the commonly used medication. In a clinical trial, 20 subjects use the new drug, and only 4 experience nausea ($\hat{\pi} = 20\%$). The research question is as follows: "**Is there evidence that patients who use the new drug are less likely to experience nausea than with the commonly used drug?**"

a. Write your null and alternative hypothesis for this investigation.

H_0: **$\pi = 0.25$ (where π represents the proportion who experience nausea after taking the new drug)**

H_a: **$\pi < 0.25$**

b. When using the binomial distribution to find the p-value, what values should you use for n and π? Keep in mind that the researchers are keeping track of the number that experience nausea.

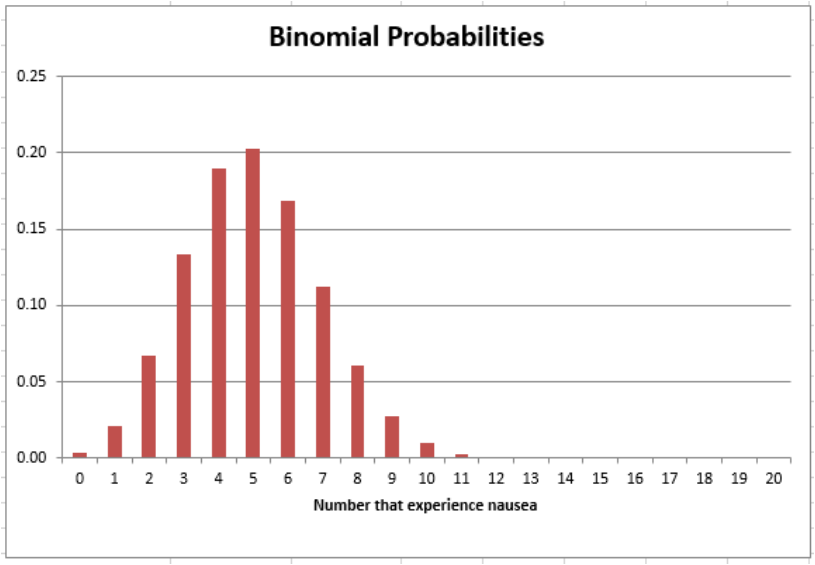
n = **20**

π = **0.25**

c. Recall that in the actual study, 4 out of 20 patients experienced nausea with the new drug. Use the binomial probabilities below to find the p-value.

p-value: **0.41**

Number that experience nausea	Binomial Probabilities
0	0.00
1	0.02
2	0.07
3	0.13
4	0.19
5	0.20
6	0.17
7	0.11
8	0.06
9	0.03
10	0.01
11	0.00
12	0.00
13	0.00
14	0.00
15	0.00
16	0.00
17	0.00
18	0.00
19	0.00
20	0.00



d. Write a conclusion in the context of the problem.

The study does not provide enough evidence to conclude that patients experience less nausea with the new drug than with the commonly used drug.

32. In the general population, the prevalence of Posttraumatic Stress Disorder (PTSD) is estimated to be about 5% in adult males. In a certain city, 40 male firefighters were randomly selected and surveyed, and it was found that 5 of the 40 ($\hat{\pi}$ = 12.5%) had PTSD. Does this result provide evidence that the prevalence of PTSD is higher for <u>all</u> male firefighters in this city than for males in the general population?

a. Write your null and alternative hypothesis for this investigation.

H_o: **$\pi = 0.05$ (where π represents the proportion of firefighters in this city who have PTSD)**

H_a: **$\pi > 0.05$**

b. When using the binomial distribution to find the p-value, what values should you use for n and π? Keep in mind that the researchers are keeping track of the number of those surveyed that have PTSD.

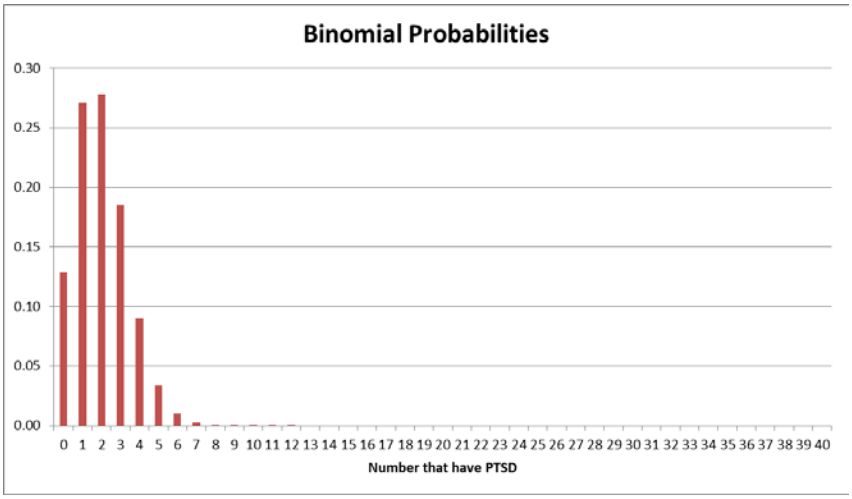
n = _____**40**_____

π = _____**0.05**_____

c. Recall that in the actual study, 5 out of 40 firefighters surveyed in this city had PTSD. Use the binomial probabilities from the next page to find the p-value.

p-value: _____**0.04**_____

Number that have PTSD	Binomial Probabilities
0	0.13
1	0.27
2	0.28
3	0.19
4	0.09
5	0.03
6	0.01
7	0.00
8	0.00
9	0.00
10	0.00
11	0.00
12	0.00
13	0.00
14	0.00
15	0.00
16	0.00
17	0.00
18	0.00
19	0.00
20	0.00
21	0.00
22	0.00
23	0.00
24	0.00
25	0.00
26	0.00
27	0.00
28	0.00
29	0.00
30	0.00
31	0.00
32	0.00
33	0.00
34	0.00
35	0.00
36	0.00
37	0.00
38	0.00
39	0.00
40	0.00



d. Write a conclusion in the context of the problem.

This result does provide evidence that the prevalence of PTSD is higher for male firefighters in this city than for males in the general population.