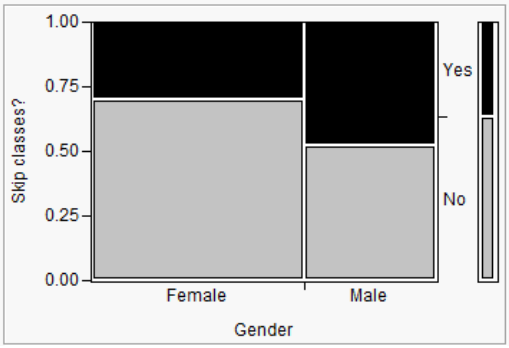
STAT 110 **Name:** _____
Exam 2 (100 pts)
Fall 2017

True/False Questions: Circle the correct answer. (2 pts each)

1.	A *risk difference* (i.e., difference in proportions) of 1 indicates no significant difference in the outcome of interest between the two groups being compared.	TRUE	(FALSE)
2.	A *relative risk ratio* (i.e., the ratio of the two risks) of 1 indicates no significant difference in the outcome of interest between the two groups being compared.	(TRUE)	FALSE
3.	Suppose a researcher believes that taking a daily vitamin C supplement will help prevent college students from catching a cold. They randomly select 300 college students and ask them about whether they take a daily vitamin C supplement and the number of colds they came down with during the winter season. This is an example of a designed experiment (i.e., a randomized controlled experiment).	TRUE	(FALSE)

Questions 4 - 9 concern a study investigating the relationship between whether a student is male or female and whether they skip class at least once a week. The following mosaic plot summarizes this relationship.



4.	There were more females surveyed than males.	(TRUE)	FALSE
5.	In the observed data, the proportion of males that skip class at least once a week is higher than that of females.	(TRUE)	FALSE
6.	Suppose the odds ratio associated with these data is 2.11, and the relative risk is 1.58. This means that the males in this study are 2.11 times as likely to skip class at least once a week as the females.	TRUE	(FALSE)
7.	If we were to carry out a hypothesis test to determine whether males were more likely to skip class, the null hypothesis would say that males and females are equally likely to skip class at least once a week.	(TRUE)	FALSE
8.	Suppose the p-value for testing the above research question was 0.01. This p-value can be interpreted as follows: If males and females in the population of all college students are equally likely to skip class at least once a week, there is only a 1% chance of observing a difference at least as big as the difference observed between males and females in this study.	(TRUE)	FALSE
9.	Consider the p-value from the previous problem. This p-value can be interpreted as follows: There is only a 1% chance that males and females are equally likely to skip class at least once a week.	TRUE	(FALSE)

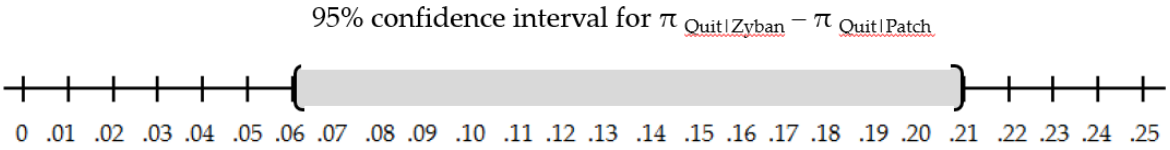
Questions 10 - 11 refer to the following scenario. Consider data that was collected on all detainees that were searched from June 2004 to June 2005 by the Los Angeles Police Department (LAPD). The table below summarizes the relationship between race of the detainee (white vs. black) and whether anything was discovered during the search (e.g., weapons, drugs, etc.).

	Discovery	No Discovery	*Total*
White	6,151	4,793	10,944
Black	10,696	13,287	23,983
Total	16,847	18,080	34,927

10. The p-value from the chi-square test of independence was found to be < 0.0001. This indicates that there is an association between race of the detainee and whether a discovery was made during the search.	(TRUE) FALSE
11. These data indicate that a discovery was more likely to be made with a black detainee than with a white detainee.	TRUE (FALSE)

Questions 12 - 14 refer to the following scenario. A study was conducted to investigate which of two treatment methods was more effective in helping people quit smoking cigarettes: a Nicotine Patch or an antidepressant medication called Zyban. Participants were randomly assigned to receive either a Nicotine Patch or the Zyban. After six months, researchers recorded whether each participant had successfully quit smoking cigarettes. A 95% confidence interval for the difference in the proportion that quit smoking was obtained and is given as follows: $0.06 \le \pi_{\text{Quit|Zyban}} - \pi_{\text{Quit|Patch}} \le 0.21$.

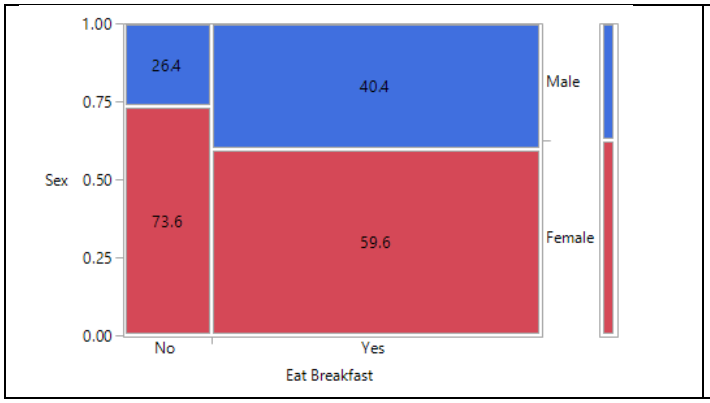
This confidence interval is also displayed in the picture below.

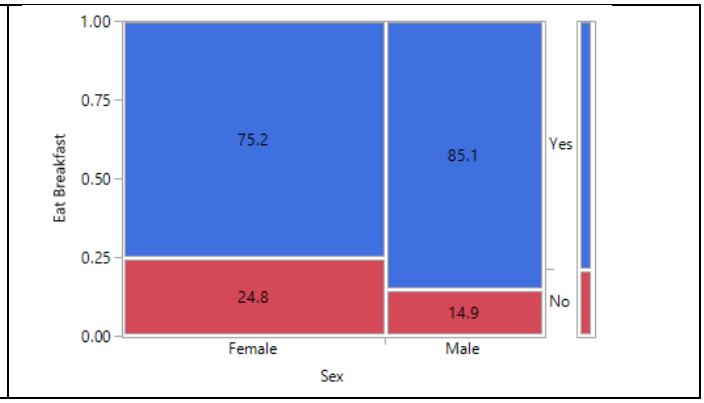


12. This confidence interval provides evidence that the proportion of people who quit smoking differs between the two treatment methods.	(TRUE) FALSE
13. The results indicate that a higher proportion of people quit smoking with the help of the nicotine patch than with the Zyban.	TRUE (FALSE)
14. The above confidence interval tells us that we can be 95% certain the proportion of people who quit smoking, overall, is between 6% and 21%.	TRUE (FALSE)

<u>Multiple Choice Questions</u>: (3 pts each)

15. A random sample of 502 college students was asked whether they eat breakfast every morning (they answered with a "yes" or a "no"). Their gender was also recorded. Two graphical representations of the results from this study are shown below.





Which of the following statements is most correct?

a. The females who were surveyed were more likely than the males who were surveyed to eat breakfast every morning.
(b.) The males who were surveyed were more likely than the females who were surveyed to eat breakfast every morning.
c. The females who were surveyed were equally likely to eat breakfast every morning as the males who were surveyed.

Questions 16 – 17 refer to the following scenario. A study was done to compare the lung capacity of coal miners to the lung capacity of farm workers. The researcher studied 200 workers of each type. Other factors that might affect lung capacity are smoking habits and exercise habits. The smoking habits of the two worker types are similar, but the coal miners generally exercise less than the farm workers.

16. Which of the following is the explanatory variable (i.e., the predictor) in this study?

a. Exercise
b. Lung capacity
c. Smoking
(d.) Occupation (coal miners vs. farm workers)

17. Which of the following is a confounding variable in this study?

(a.) Exercise
b. Lung capacity
c. Smoking
d. Occupation (coal miners vs. farm workers)

Questions 18 – 19 refer to the following scenario. Researchers randomly selected participants and then randomly assigned them to either walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. At the end of a year, the change in the participants' blood pressure over the year was measured, and the change was compared for the two groups.

18. This is a randomized controlled experiment (i.e., a designed experiment) because

a. Blood pressure was measured at the beginning and end of the study.
b. The two groups were compared at the end of the study.
(c.) The participants were randomly assigned to either walk or read, rather than choosing their own activity.
d. A random sample of participants was used.

19. If a statistically significant difference in blood pressure change between the two activities is found, then which of the following conclusions is most correct?

a. It cannot be concluded that the difference in activity caused a difference in the change in blood pressure because in the course of a year there are lots of possible confounding variables.
b. It cannot be concluded that the difference in activity caused a difference in the change in blood pressure because it might be the opposite; people with high blood pressure might be more likely to read a book than to walk.
(c.) It can be concluded that the difference in activity caused a difference in the change in blood pressure because of the way the study was done; the randomization of subjects to treatment groups should balance out the effects of any confounding variables.

Questions 20 – 21 refer to the following scenario. To assess the effects of tanning beds on the incidence of skin cancer, a researcher collected and compared mole biopsies from individuals who reported that they used tanning beds regularly and compared them to the mole biopsies from individuals who reported never using a tanning bed. The researcher analyzed the moles for cancerous cells and was unaware of which samples were from the regular tanning group and the group that had never used a tanning bed.

20. Is this an observational study or a designed experiment?

(a.) Observational study because the people themselves chose whether to use tanning beds.
b. Designed experiment (i.e., randomized controlled experiment) because one group used tanning beds and the other group did not.
c. There is not enough information provided to decide whether this was an observational study or a designed experiment.

21. Suppose there were a significantly higher number of cancerous moles for the tanning bed users than for the non-tanning bed users. Can you conclude based on this study that using tanning beds leads to cancerous moles?

a. Yes - the researcher was blind to which samples belonged to those who used tanning beds and those who didn't.
(b.) No - there may be other lifestyle differences between people who choose to tan and people who do not (e.g., sunscreen habits, differences in diet, etc.)
c. Yes - there were a significantly higher number of cancerous moles in the individuals who used tanning beds.

Questions 22 – 24 refer to the following scenario. Suppose that a survey study was used to classify subjects according to both their gender and whether they owned a pet. Both the observed counts and the expected counts needed for a chi-square test are shown below.

Observed Counts		
	Has a Pet	No Pet
Male	8	22
Female	16	14

Expected Counts		
	Has a Pet	No Pet
Male	12	18
Female	12	18

22. Which of the following shows the correct calculations for the chi-square test statistic?

a. $\frac{(8-12)^2}{12}+\frac{(22-18)^2}{18}$

b. $\frac{(16-12)^2}{12}+\frac{(14-18)^2}{18}$

c. $1.96\sqrt{\frac{8/30(1-8/30)}{30}+\frac{16/30(1-16/30)}{30}}$

(d.) $\frac{(8-12)^2}{12}+\frac{(22-18)^2}{18}+\frac{(16-12)^2}{12}+\frac{(14-18)^2}{18}$

23. If the p-value associated with the chi-square test statistic for this problem is less than 0.05, then which of the following conclusions is most correct?

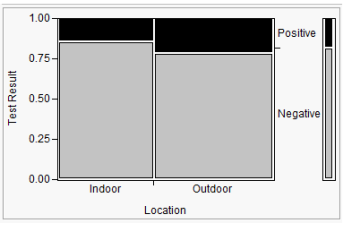
a. There is not enough evidence to conclude that an association exists between gender and pet ownership.
b. There is a significant difference between the sample and the population.
c. There is no significant association between the sample and the population.
(d.) There is evidence that an association exists between gender and pet ownership.

24. If the p-value associated with the chi-square test statistic for this problem is less than 0.05, then which of the following statements is most correct?

(a.) It would be very surprising to obtain the observed sample results if there is really no association between gender and pet ownership.
b. It would be very surprising to obtain the observed sample results if there is really an association between gender and pet ownership.
c. It would not be surprising to obtain the observed sample results if there is really no association between gender and pet ownership.
d. It would not be surprising to obtain the observed sample results if there is really an association between gender and pet ownership.

25. To investigate whether there is a relationship between where a dog is kept (indoors or outdoors) and the dog's risk of acquiring Lyme's disease, 94 dogs were examined and classified according to their location (indoors or outdoors) and the result of their Lyme's disease test (positive or negative). The data were analyzed in JMP, and the results are shown below.

	Test Result		
Locatio	Count	Negative	Positive	
	Indoor	36	6	42
	Outdoor	41	11	52
		77	17	94



Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.8712	Prob(Test Result=Positive) is greater for Location=Indoor than Outdoor
Right	0.2793	Prob(Test Result=Positive) is greater for Location=Outdoor than Indoor
2-Tail	0.4317	Prob(Test Result=Positive) is different across Location

Which of the following sets of statements follows from these results?

a. The study provides evidence to support the theory that dogs which are kept indoors have a higher chance of acquiring Lyme's disease.
b. The study provides evidence to support the theory that dogs which are kept outdoors have a higher chance of acquiring Lyme's disease.
(c.) The study does not provide enough evidence to show that acquiring Lyme's disease is associated with where a dog is kept (i.e., there is not enough evidence to show that the proportion testing positive differs between the indoor and outdoor dogs).

Short Answer Questions:

26. A study was conducted to examine the association between the type of crowd high school students typically associated with and the ways they volunteered in their community. The data are summarized in the following table.

	Voluntary Service Categories			
Crowd Type	Helping People in Need	Working with Children	No Service	Totals
School (e.g., study groups, clubs, etc.)	38	30	23	91
Fun (e.g., party crowds)	13	22	61	96
All-around (both school and fun crowds)	21	23	37	81
Disengaged	19	24	50	93
Totals	91	99	171	361

a. Use relevant percentages to explain how the proportion that do No Service in their community differs across the four levels of crowd type in this data set. Be sure to indicate which crowd types are most (and least) likely to do no service, and report the relevant percentages. (4 pts)

Students who were part of the Fun crowd were most likely to not serve the community in any way (63.5% said they did no service). Of those in the Disengaged crowd, 53.8% did no service; 45.7% of the All-Around crowd did no service. Students in the School Crowd were the least likely to not serve the community in any way (only 25.3% did no service).

b. Set up the null and alternative hypothesis to test whether there is an association between Crowd Type and Voluntary Service Category. (3 pts)

H_o: **There is no association between Crowd Type and Voluntary Service Category.**

H_a: **There is an association between Crowd Type and Voluntary Service Category.**

c. Compute the two expected counts highlighted in the table below (note that these would be used to compute the chi-square test statistic). (2 pts)

	Helping People in Need	Working with Children	No Service	Totals
School (e.g., study groups, clubs, etc.)				91
Fun (e.g., parties)			**45.47**	96
All-around (both school and fun crowds)		**22.2**		81
Disengaged				93
Totals	91	99	171	361

d. Identify the value of the chi-square statistic from the JMP output.

Tests

N	DF	-LogLike	RSquare (U)
361	6	17.072846	0.0448

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	34.146	<.0001*
Pearson	33.308	<.0001*

Chi-square test statistic: **33.308** (2 pts)

e. Identify the p-value associated with the chi-square statistic from the JMP output.

p-value: **<0.0001** (3 pts)

f. Write a conclusion in the context of the problem. (3 pts)

There is evidence that an association exists between Crowd Type and Voluntary Service Category.

27. Suppose a study was conducted to investigate whether the rate of injury due to alcohol differs across gender in college students identified with at-risk drinking behavior. The study screened all students using a university's health center and identified those with at-risk drinking behavior; 2,090 of these students agreed to participate in this study and were classified according to their gender and whether they had suffered any alcohol-related injuries in the past six months. The data are summarized in the following table.

	Injury Due to Alcohol?		
Gender	Yes	No	*Totals*
Male	101	771	872
Female	110	1,108	1,218
Totals	211	1,879	2,090

a. Find the risk of an alcohol-related injury for males. (2 pts)

101/872 = 0.1158

b. Find the risk of an alcohol-related injury for females. (2 pts)

110/1218 = 0.0903

c. Find the relative risk ratio (also known as the relative risk). (2 pts)

0.1158/0.0903 = 1.282

d. Interpret the relative risk ratio from part c. (2 pts)

Males were 1.282 times as likely to suffer an alcohol-related injury as were females in this study.

e. Find the odds of an alcohol-related injury for males. (2 pts)

101/771

f. Find the odds of an alcohol-related injury for females. (2 pts)

110/1108

g. Find the odds ratio. (2 pts)

(101/771) / (110/1108) = 1.32

h. Interpret the odds ratio from part g. (2 pts)

The odds of an alcohol-related injury were 1.32 times as large for males as for females in this study.

28. Consider a study in which the relative risk ratios associated with various risk factors for persistent post-partum depression (PPD) were being considered (persistent means that symptoms lasted up to one year after giving birth).

Risk factor for Persistent PPD	Risk Ratio (i.e., Relative Risk)
≤ 20 years of age	0.9
Pregnant for the First Time	0.6
Husband Not Educated	1.7
Not Educated	1.2
Poverty	1.4

For example, the first risk ratio was computed as follows:

$$\text{Risk Ratio} = \frac{Risk\ of\ Persistent\ PPD\ for\ Women\ Age \leq 20}{Risk\ of\ Persistent\ PPD\ for\ Women\ Age > 20} = 0.9$$

a. Which group was more at risk for PPD? Circle your answer. (1 pt)

Women Age ≤ 20 (Women Age > 20)

b. Explain the reasoning behind your answer to part a. (1 pt)

Since the relative risk ratio was less than one, the risk was larger for the group whose risk was used in the denominator.

c. Of the risk factors listed in the table, which appears to be the *most* important in determining if a woman will experience persistent post-partum depression? Explain your reasoning. (2 pts)

Whether the husband was educated, since this risk ratio was farthest away from one.

d. Of the risk factors listed, which appears to be the *least* important in determining if a woman will experience persistent post-partum depression? Explain your reasoning. (2 pts)

Whether the mother was ≤ 20 years of age, since this risk ratio was closest to one.