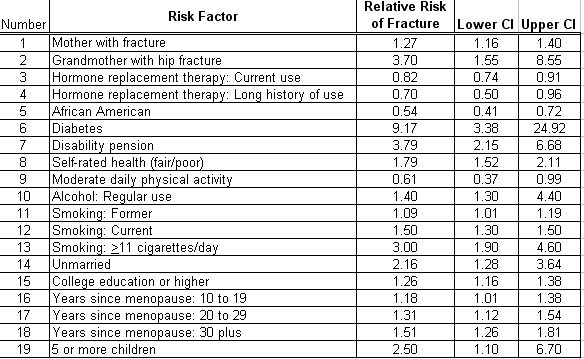
STAT 210: Exam #2 Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Consider the following study in which various risk factors were being considered as a method for screening for postmenopausal osteoporosis.  
  
Source: <http://www.ahrq.gov/clinic/3rduspstf/osteoporosis/osteosumm1.htm>  
  
The risk factors under consideration have been numbered from 1 – 19 in the following table.



As an example, consider the 1st Risk Fracture (Mother with Fracture). The reported relative risk ratio was computed as follows:

1. Answer the following
2. Consider the relative risk for Diabetes at 9.17. Using everyday language, explain what this value means?
3. Look at Risk Factors #11, #12, and #13. What can be said about the effects of smoking in relation to bone fractures in females? Discuss.
4. The risk factors listed above were those found to be statistically important. In particular, notice that none of the confidence intervals capture 1.0. Why is it the case that none of the confidence intervals contain 1.0? In particular, what would it mean if the confidence interval did contain 1.0? Explain.

The September 2013 issue of *Pediatrics* reported a study involving 1,232 adolescents. They were classified according to whether or not they were adopted and whether or not they had attempted suicide. The data are summarized in the following contingency table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Attempted Suicide | Did Not Attempt Suicide | Total |
| Adopted | 47 | 645 | 692 |
| Not Adopted | 9 | 531 | 540 |
| Total | 56 | 1176 | 1232 |

1. Suppose you were to find the relative risk for these data as follows:

What is this relative risk ratio?

* 1. (47/645) / (9/531) = 4.30
  2. (47/56) / (9/56) = 5.22
  3. (47/692) / (9/540) = 4.08
  4. (47/1232) / (9/1232) = 5.22

1. Provide the name for the statistical quantity that would be used to fill in the blank in the following sentence? *“An adolescent in this study who was adopted is \_\_\_\_\_ times more likely to attempt suicide than an adolescent in this study who was not adopted.”*
2. risk difference
3. relative risk ratio
4. odds ratio
5. Suppose you were to find the odds ratio for these data as follows:

What is this odds ratio?

1. (47/645) / (9/531) = 4.30
2. (47/56) / (9/56) = 5.22
3. (47/692) / (9/540) = 4.08
4. (47/1232) / (9/1232) = 5.22

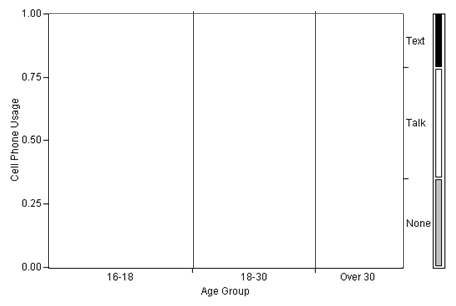
Consider the following data on the investigation of Age of Driver and Cell Phone Usage for car accidents.

Research Question: Does age have an influence on the type of cell phone usage of drivers involved in car accident?

|  |  |
| --- | --- |
|  |  |

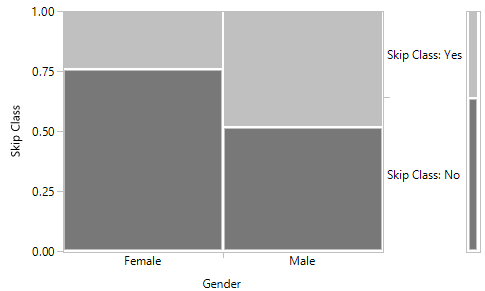
Answer the following using the above JMP output.

1. What is the p-value for this test? \_\_\_\_\_\_\_\_\_\_\_
2. Which of the following is the best conclusion for this research question?
3. The data supports the research question because the p-value is less than 0.05.
4. We have evidence to suggest that Age Group influences the type of cell phone usage of drivers involved in a car accident because the p-value is less than 0.05.
5. The patterns in the graph are different which implies that Age Group influences cell phone usage.
6. Sketch a different mosaic plot that would provide even more evidence that Age Group influences cell phone usage. Sketch you graph carefully and using the same color scheme as above (Text = Black, Talk = White, and None=Gray).



Consider a study investigating whether a relationship exists for WSU students between a student’s gender and whether they skip class at least once a week. The following mosaic plot summarizes the data that was collected.

Research Question: Are Males more likely than Females to skip class at least once a week?



1. Answer the following True/False Questions

|  |  |  |
| --- | --- | --- |
| 1. We know that in this study data was collected on more females than males because WSU has a lot more females on its campus. | TRUE | FALSE |
| 1. For this data, the percentage of males that skip class at least once a week is higher than that of females. | TRUE | FALSE |
| 1. Suppose the odds ratio associated with these data is 2.11, and the relative risk of skipping class is 1.58 (males were used in the numerator of both). This means that the males are 2.11 times more likely to skip class at least once a week than females. | TRUE | FALSE |
| 1. If we were to carry out a statistical test to determine whether males were more likely to skip class, the expected counts would be setup up under that situation that males and females are equally likely to skip class at least once a week. | TRUE | FALSE |
| 1. Suppose the p-value for testing the above research question was 0.01 (or 1%). This p-value can be interpreted as follows: There is a 1% chance of getting a difference as extreme or more extreme than the one observed under the situation that females and males are equally likely to skip class at least once a week. | TRUE | FALSE |
| 1. Consider again the p-value of 0.01 (or 1%). 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 |
| 1. Consider the p-value = 0.01 (or 1%). This study provides enough statistical evidence that males are more likely to skip class at least once a week than are females (p-value = 0.01). | TRUE | FALSE |

A study investigated whether 3-year-old, 4-year-old, and 5-year-old children learned from others’ conversations even when they had not been instructed to pay attention to the conversation and when they were engaged in another activity. Four different colored drawers (red, blue, yellow, and orange) were placed in the test area, and the experiment was carried out separately for each child. They placed the child off to the side, distracted them with a toy, and had a scripted conversation while the child played. Towards the end of the scripted conversation, one of the researchers stated that she was going to check to make sure the prize was ready. Then she said, ‘‘Oh, wait, I don’t remember which drawer has the prize in it. Is it the (red) drawer?’’ The other researcher responded, ‘‘Yes, it’s the (red) one.’’ A few seconds later, when they had brought the child back into the conversation, they asked the child to identify which drawer contained the prize and recorded whether or not the child chose the correct drawer. Note that the color of the drawer that the toy was in was counterbalanced between participants.

When analyzing the results, the researchers considered both age of the child and whether or not they chose the correct drawer. The results from JMP were as follows:

|  |  |
| --- | --- |
|  | Table of Counts    Chi-Square Test |

Research Question: Is there statistical evidence to say the likelihood of picking the correct drawer is different across Age (3, 4, and 5 Year Olds)?

1. Answer the following True/False Questions

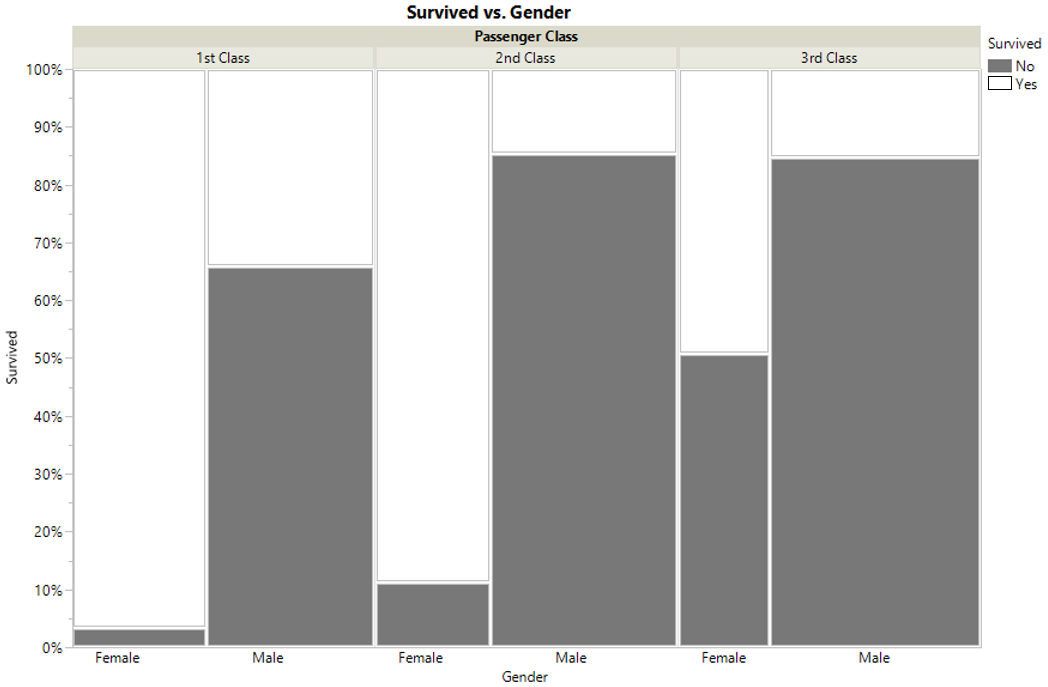
|  |  |
| --- | --- |
| 1. In order to investigate whether the proportion choosing the correct drawer differs across age group, it is most appropriate to compare the proportions 11/33 to 10/33 to 12/33. | TRUE FALSE |
| 1. In order to investigate whether the proportion choosing the correct drawer differs across age group, it is most appropriate to compare the proportions 11/16 to 10/17 to 12/18. | TRUE FALSE |
| 1. We have statistical evidence that the proportion choosing the correct drawer differs across age group simply because the mosaic plot shows us that more 4-year-olds in this study chose the wrong drawer. | TRUE FALSE |
| 1. The p-value of 0.8177 indicates that this study does not provide enough evidence that the proportion choosing the correct drawer differs across age group. | TRUE FALSE |
| 1. The p-value of 0.8177 can be used to test whether or not children overall (regardless of age) are choosing the correct drawer more often than we’d expect by chance. | TRUE FALSE |

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 Type (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.

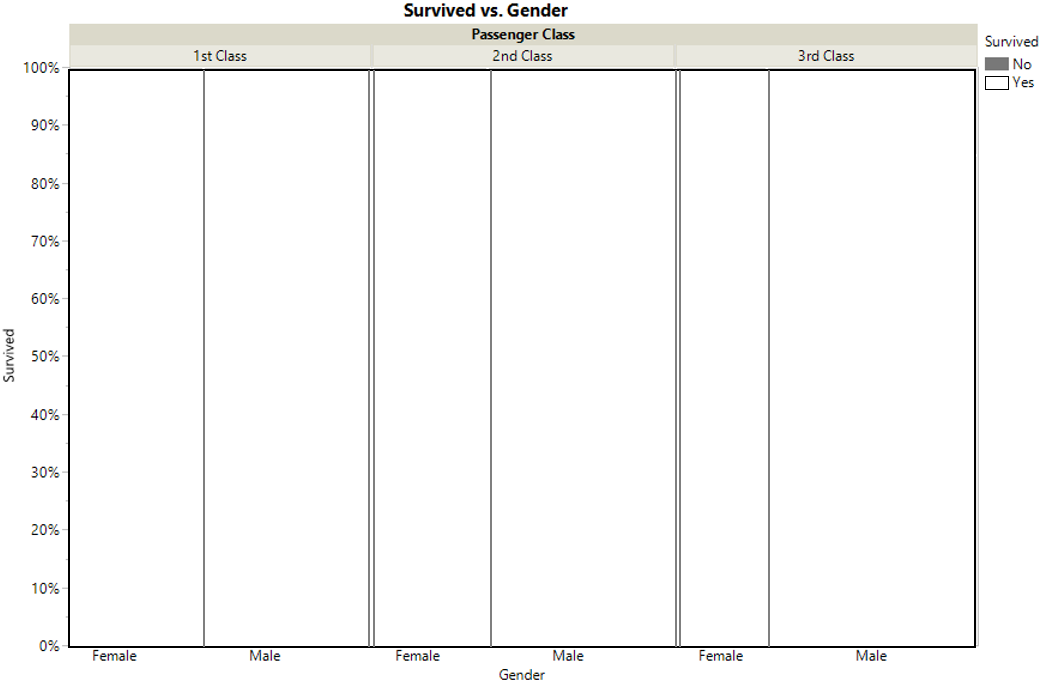
|  |  |
| --- | --- |
|  |  |

1. Which of the following sets of statements follows from these results?
   1. The study does not provide enough statistical 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).
   2. The study provides statistical evidence to support the theory that dogs which are kept indoors have a higher chance of acquiring Lyme’s disease.
   3. The study provides statistical evidence to support the theory that dogs which are kept outdoors have a higher chance of acquiring Lyme’s disease.
2. Compute the following quantities and interpret each.
   1. Compute the relative risk difference for dogs testing positive:
   2. Interpret the relative risk difference in the context of this problem:
   3. Compute the relative risk ratio for dogs who test positive:
   4. Interpret the relative risk ratio in the context of this problem:

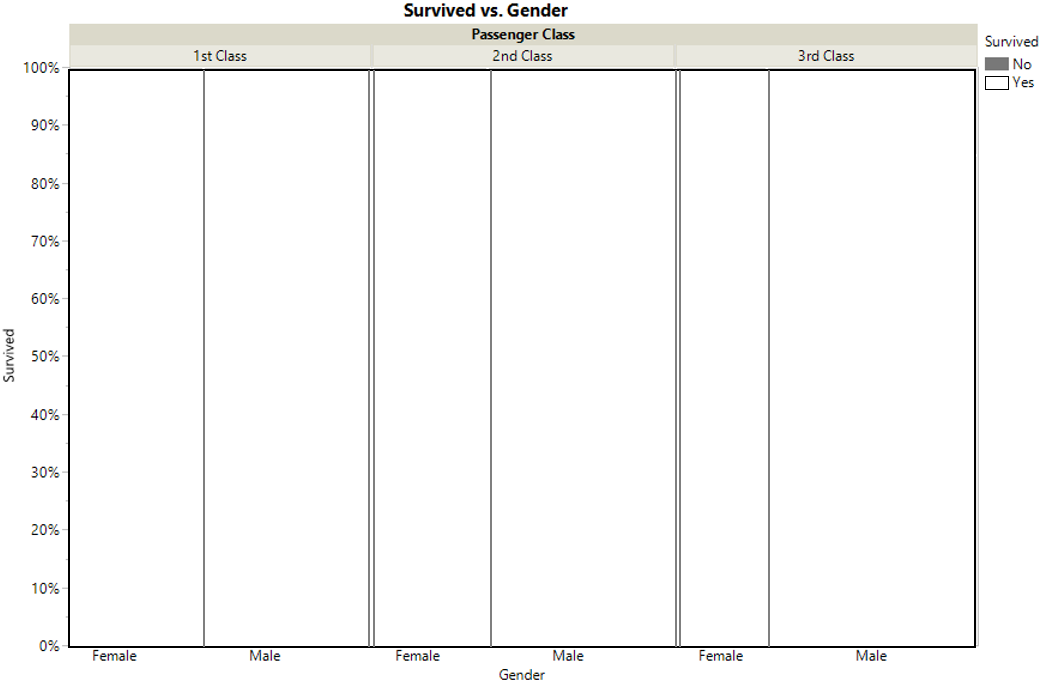
Consider the following association between Gender and Passenger Class on the Survival Rate of those on the Titanic.



1. Give a sketch below that would clearly exhibit the following behavior: There is no association between Survival Rate and Gender *and* no association between Survival Rate and Passenger Class.



1. Give a sketch below that would clearly exhibit the following behavior: Gender has an association with Survival Rate, but Passenger Class does not.



**EXTRA MATERIAL FROM HANDOUT #9** (Not covered on their Exam #2)

1. Answer the following True/False questions.

|  |  |  |  |
| --- | --- | --- | --- |
| a. | The average is measured on the same scale as the data. That is, if the data is in dollars, then the average is expressed in dollars as well. | TRUE | FALSE |
| b. | The standard deviation is measured on the same scale as the data. That is, if the data is in dollars, then the standard deviation is expressed in dollars as well. | TRUE | FALSE |
| c. | The range is the most widely accepted method of measuring spread for a set of data by statisticians. | TRUE | FALSE |
| d. | The standard deviation uses all the data points in its calculation. | TRUE | FALSE |
| e. | Z-scores (also known as standardized values) are used to identify outliers in a set of data. | TRUE | FALSE |
| f. | If a data point has a Z-Score below -2 or above +2, then it is considered an outlier. | TRUE | FALSE |

Consider location across Minnesota for which snowfall amounts have been collected. For this problem, you can assume that the further north the location, the more snow that location gets.

|  |  |  |
| --- | --- | --- |
| Original Data | New Data #1 | New Data #2 |
|  |  |  |

1. Answer the following True/False questions. (1 pt each)

|  |  |  |  |
| --- | --- | --- | --- |
| a. | We would expect that New Data #2 would have the largest average snowfall amount. | TRUE | FALSE |
| b. | We would expect that the Original Data would have the smallest standard deviation because it has the fewest number of locations. | TRUE | FALSE |
| c. | We would expect the standard deviation to be about the same between the Original Data and New Data#1. | TRUE | FALSE |
| d. | We would expect that the standard deviation for New Data #2 to be smaller than the standard deviation from New Data #1. | TRUE | FALSE |