Study Guide for Exam #2
Fall 2017

STAT 210: Exam #2 Name: SOLUTION

1. 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.

Number	Risk Factor	Relative Risk of Fracture	Lower CI	Upper CI
1	Mother with fracture	1.27	1.16	1.40
2	Grandmother with hip fracture	3.70	1.55	8.55
3	Hormone replacement therapy: Current use	0.82	0.74	0.91
4	Hormone replacement therapy: Long history of use	0.70	0.50	0.96
5	African American	0.54	0.41	0.72
6	Diabetes	9.17	3.38	24.92
7	Disability pension	3.79	2.15	6.68
8	Self-rated health (fair/poor)	1.79	1.52	2.11
9	Moderate daily physical activity	0.61	0.37	0.99
10	Alcohol: Regular use	1.40	1.30	4.40
11	Smoking: Former	1.09	1.01	1.19
12	Smoking: Current	1.50	1.30	1.50
13	Smoking: ≥11 cigarettes/day	3.00	1.90	4.60
14	Unmarried	2.16	1.28	3.64
15	College education or higher	1.26	1.16	1.38
16	Years since menopause: 10 to 19	1.18	1.01	1.38
17	Years since menopause: 20 to 29	1.31	1.12	1.54
18	Years since menopause: 30 plus	1.51	1.26	1.81
19	5 or more children	2.50	1.10	6.70

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

$$\text{Relative Risk of Fracture} = \frac{\%\text{ Fracture for those who's Mother has had a bone fracture}}{\%\text{ Fracture for those who's Mother has NOT had a bone fracture}} = 1.27$$

Answer the following

a. Consider the relative risk for Diabetes at 9.17. Using everyday language, explain what this value means?

   A woman with Diabetes is 9.17 times more likely to have a bone fracture than a women who does not have diabetes.

b. 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.

   First, a former smokers bone density is about the same as a nonsmokers, relative risk ratio pretty close to 1 (although technically different than 1 as the CI does not capture 1 here). Risk Factor #13 suggest the more you smoke the more

your risk for fracture increases and is about 3 times higher.

c. 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? What would it mean if the confidence interval did contain 1.0? Explain.

If the relative risk factor is 1 then there is no difference in risk. If risk factors are statistically important, then the 95% CI will not capture 1. [Recall the 95% confidence interval is simple the estimate +/- margin-of-error, this is used to obtain a lower and upper endpoint.] If the interval did happen to capture 1, then the p-value from the statistical test would be above 0.05, i.e. *not* statistically significant.

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

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

$$Risk\ Ratio = \frac{Risk\ of\ attempting\ suicide\ if\ adopted}{Risk\ of\ attempting\ suicide\ if\ not\ adopted}$$

What is this relative risk ratio?

a. (47/645) / (9/531) = 4.30
b. (47/56) / (9/56) = 5.22
c. (47/692) / (9/540) = 4.08
d. (47/1232) / (9/1232) = 5.22

3. 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."*

a. risk difference
b. relative risk ratio
c. odds ratio

4. Suppose you were to find the odds ratio for these data as follows:

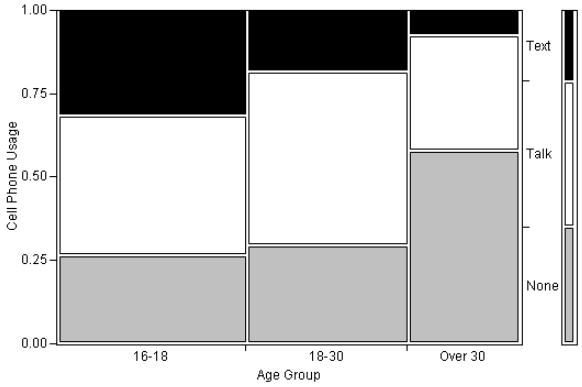
$$Odds\ Ratio = \frac{Odds\ of\ attempting\ suicide\ if\ adopted}{Odds\ of\ attempting\ suicide\ if\ not\ adopted}$$

What is this odds ratio?

(a.) (47/645) / (9/531) = 4.30
b. (47/56) / (9/56) = 5.22
c. (47/692) / (9/540) = 4.08
d. (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?



Tests

N	DF	-LogLike	RSquare (U)
388	4	19.608793	0.0478

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	39.218	<.0001*
Pearson	39.470	<.0001*

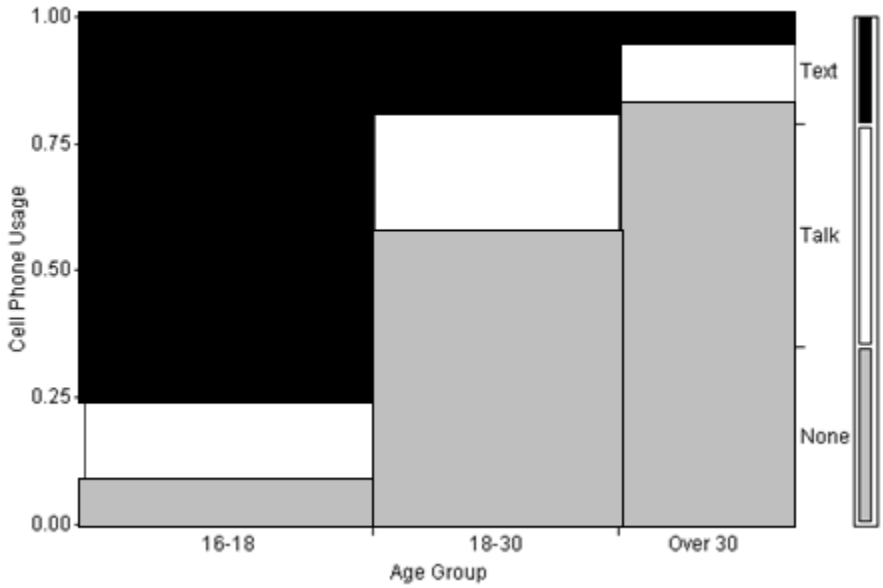
Answer the following using the above JMP output.

5. What is the p-value for this test? < 0.0001

6. Which of the following is the best conclusion for this research question?

a. The data supports the research question because the p-value is less than 0.05.
(b.) 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.
c. The patterns in the graph are different which implies that Age Group influences cell phone usage.

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).



Concern 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?

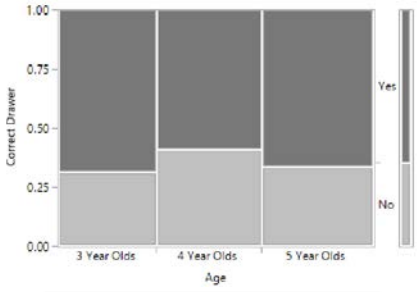
7. Answer the following True/False Questions

a. 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)
b. For this data, the percentage of males that skip class at least once a week is higher than that of females.	(TRUE) FALSE
c. 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)
d. 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
e. 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
f. 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)
g. 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 Excel were as follows:

Graph & P-Value



Tests

N	DF	-LogLike	RSquare (U)
51	2	0.19962420	0.0060

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.399	0.8190
Pearson	0.402	0.8177

Table

Contingency Table

Count	Correct Drawer: No	Yes	Total
Age: 3 Year Olds	5	11	16
4 Year Olds	7	10	17
5 Year Olds	6	12	18
Total	18	33	51

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)?

8. Answer the following True/False Questions

a. 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)
b. 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
c. 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)
d. 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
e. 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 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 location (indoors or outdoors) and the result of their Lyme's disease test (positive or negative). The data were analyzed in Excel, and the results are shown below.

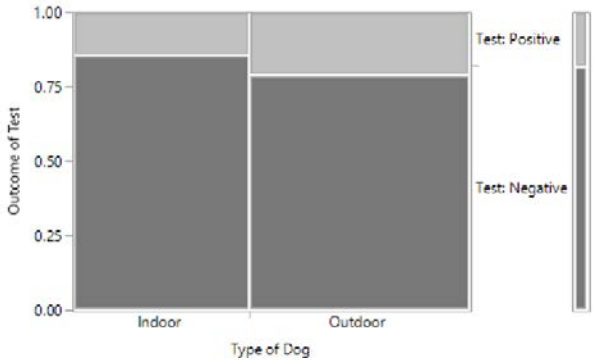
Contingency Table

		Outcome of Test		
	Count Row %	Test: Negative	Test: Positive	Total
Type of Dog	Indoor	36 85.71	6 14.29	42
	Outdoor	41 78.85	11 21.15	52
	Total	77	17	94

Tests

N	DF	-LogLike	RSquare (U)
94	1	0.37580920	0.0085

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	0.752	0.3860
Pearson	0.740	0.3897



9. Which of the following sets of statements follows from these results?
   a. 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).
   b. The study provides statistical evidence to support the theory that dogs which are kept indoors have a higher chance of acquiring Lyme's disease.
   c. The study provides statistical evidence to support the theory that dogs which are kept outdoors have a higher chance of acquiring Lyme's disease.

10. Compute the following quantities and interpret each.

    a. Compute the relative risk difference for dogs testing positive: $21.15\% - 14.29\% = 6.86\% \approx 7\%$

    b. Interpret the relative risk difference in the context of this problem:

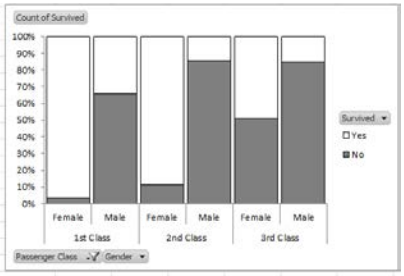
    The chance of a dog testing positive for Lyme's disease is about 7% higher for an Outdoor dog compared to an Indoor dog.

    c. Compute the relative risk ratio for dogs who test positive: $\frac{21.15\%}{14.29\%} = 1.48 \approx 1.5$

    d. Interpret the relative risk ratio in the context of this problem:

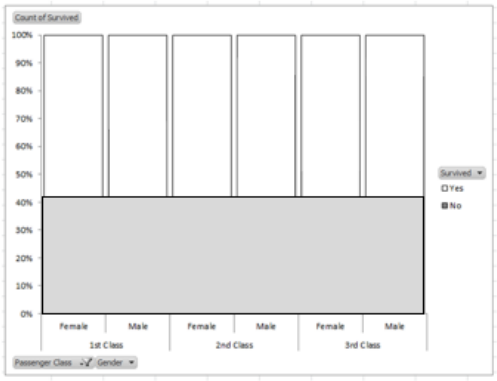
    An Outdoor dog is about 1.5 times more likely to test positive for Lyme's disease than an Indoor dog.

Consider the possible association between Gender and Passenger Class on the Survival Rate of those on the Titanic. The graph from the Titanic is provided here. On the graphs below, give a rough sketch for each situation.



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

No differences at all…



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

Differences between Genders, but not Passenger classes…