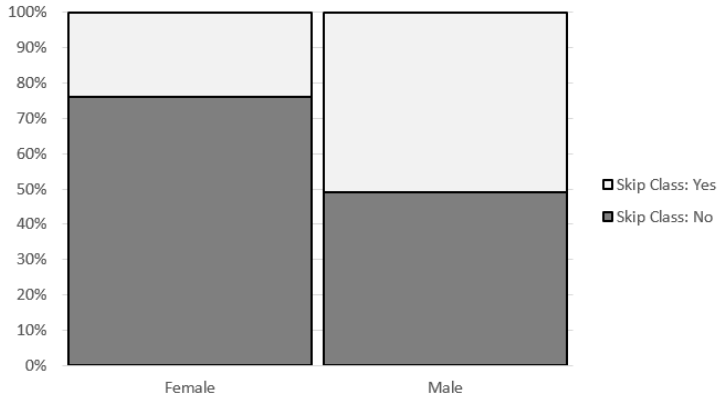
STAT 110: Exam #2 Name: SOLUTION

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.



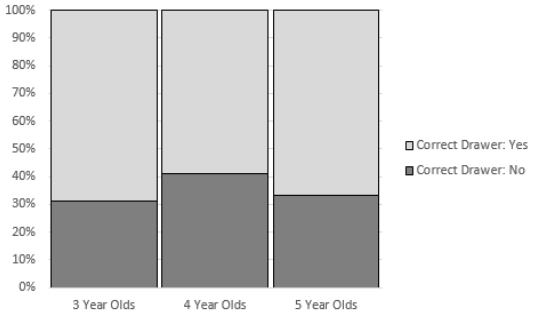

Research Question: Are Males more likely than Females to skip class at least once a week?

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) |
| 2. For this data, the percentage of males that skip class at least once a week is higher than that of females. | (TRUE) FALSE |
| 3. 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) |
| 4. 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 |
| 5. 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 |
| 6. 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) |
| 7. 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:




| | A | B | C | D |
|---|---|---|---|---|
| 1 | Observed | | | |
| 2 | Correct Drawer | No | Yes | Total |
| 3 | 3 Year Olds | 5 | 11 | 16 |
| 4 | 4 Year Olds | 7 | 10 | 17 |
| 5 | 5 Year Olds | 6 | 12 | 18 |
| 6 | Total | 18 | 33 | 51 |
| 7 | | | | |
| 8 | Expected | | | |
| 9 | Correct Drawer | No | Yes | Total |
| 10 | 3 Year Olds | 5.6 | 10.4 | 16 |
| 11 | 4 Year Olds | 6 | 11 | 17 |
| 12 | 5 Year Olds | 6.4 | 11.6 | 18 |
| 13 | Total | 18 | 33 | 51 |
| 14 | | | | |
| 15 | | =CHITEST(B3:C5,B10:C12) | | |
| 16 | P-Value | 0.8177 | | |

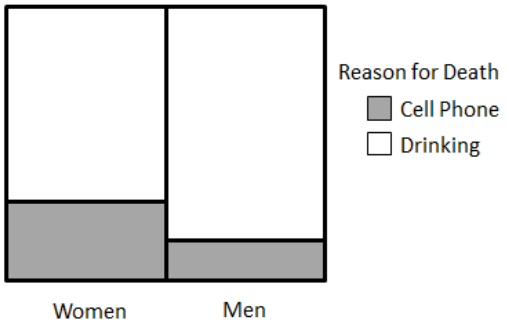
Research Question: Is there statistical evidence to say the likelihood of picking the correct drawer is different across Age (3, 4, and 5 Year Olds)?

Answer the following True/False Questions

| | |
|---|---|
| 8. 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) |
| 9. 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 |
| 10. 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) |
| 11. 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 |
| 12. 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) |

Consider the following graph that shows the relationship between Gender and Reason for Death in fatal car accidents for people whose age is between 18 and 24. This study included deaths from car accidents for 100 women and 200 males.

Research Question: Do difference exist in the risk of death due to Cell Phones between Women and Men?




13. Consider the following rationale as to why a statistician would want to run a formal statistical test. Which of the following rationale is correct?

    a. A statistical test will allow us to generalize the outcomes from this study to other people age 18-24 that are not necessarily in this study.
    b. A statistical test will permit us to take into consideration that this graph may change over repeated samples.
    c. A statistical test is only necessary when the number of women and men are different in the study.
    (d.) The rationale for a. and b. is correct
    e. The rationale for all three is correct

Suppose we used data collected from a survey of Winona State students to test for a relationship between gender and alcohol consumption. The observed counts are shown in the following contingency table.

| | Consumes Alcohol? | | |
|---|---|---|---|
| | No | Yes | Totals |
| Female | 101 | 101 | 202 |
| Male | 35 | 111 | 146 |
| Totals | 136 | 212 | 348 |

14. If we were to find each of the expected counts needed to compute a chi-square test statistic, which of the following would be the expected number of males who do consume alcohol?

    a. 106
    (b.) 88.94
    c. 111
    d. 87

15. The p-value for the previous study comparing alcohol assumption across gender is < 0.0001. Which of the following is a valid conclusion?

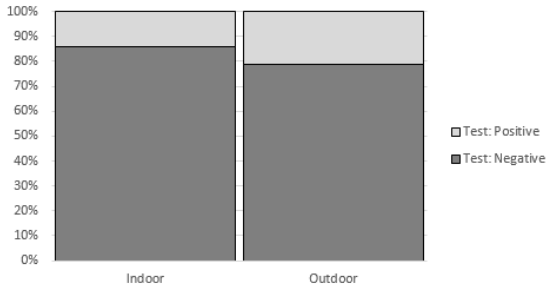
    (a.) It would be very surprising to obtain the observed sample results if there is really no difference between men and women.
    b. It would not be surprising to obtain the observed sample results if there is really no difference between men and women.
    c. It would be very surprising to obtain the observed sample results if there is really a difference between men and women.

16. To test the effectiveness of a new drug for high blood pressure, 20 subjects with high blood pressure were randomly divided into two groups. Group 1 received the medication, and Group 2 received a placebo. A statistical test was used to test whether the proportion that found relief from their symptoms was significantly higher for Group 1. The p-value from this test was 0.001. Which of the following statements is most correct?

    a. There is no evidence that the medication provides relief from their high blood pressure symptoms.
    b. There is enough evidence that the medication provides relief from high blood pressure symptoms.
    c. There is enough evidence that the medication provides relief form high blood pressure symptoms; however, the reason for the relief is likely due to other factors that are not being considered in this study.
    d. The sample size is too small to draw a valid conclusion.

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.

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Observed | | | |
| 2 | | Test: Negative | Test: Positive | Total |
| 3 | Indoor | 36 | 6 | 42 |
| 4 | Outdoor | 41 | 11 | 52 |
| 5 | Total | 77 | 17 | 94 |
| 6 | | | | |
| 7 | Expected | | | |
| 8 | | Test: Negative | Test: Positive | Total |
| 9 | Indoor | 34.4 | 7.6 | 42 |
| 10 | Outdoor | 42.6 | 9.4 | 52 |
| 11 | Total | 77 | 17 | 94 |
| 12 | | | | |
| 13 | | =CHITEST(B3:C4,B9:C10) | | |
| 14 | P-Value | 0.3897 | | |




17. 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.

18. Compute the following quantities and interpret each.

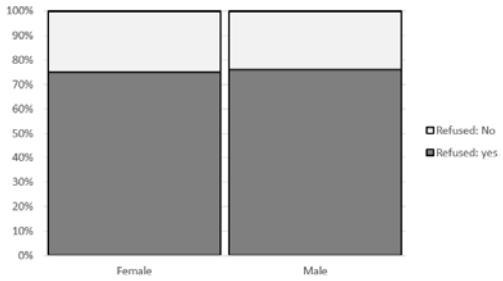
$$\text{Risk of Positive test for Outdoor dogs: } \frac{11}{52} \approx 0.2115 = 21.15\%$$
$$\text{Risk of Positive test for Indoor dogs: } \frac{6}{42} \approx 0.1429 = 14.29\%$$

    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 risk that an outdoor dog will test positive for Lyme's disease is about 7% higher than the risk of an indoor dog testing positive.

    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:
    The risk that an outdoor dog will test positive for Lyme's disease is about 1.5 times higher than the risk of an indoor dog testing positive for Lyme's disease.

In a survey of families in which both parents work, one of the questions asked was, “Have you refused a job, promotion, or transfer because it would mean less time with your family?" About 250 men and 500 women were asked this question. Consider the following hypotheses and mosaic plot.

Research Question: Are there differences in the rate of refusal between Females and Males?




19. The p-value for this analysis is most likely to be which of the following, and why?
    a. Less than 0.05 because there are approximately twice as many females as males.
    b. Greater than 0.05 because there are approximately twice as many females as males.
    c. Less than 0.05 because the patterns for each row in the mosaic plot are similar across the columns.
    (d.) Greater than 0.05 because the patterns for each row in the mosaic plot are similar across the columns.

20. The General Social Survey is a massive survey of American adults that is conducted yearly. Sometimes, the researchers ask a question in two different forms (Form X and Form Y). About half the people surveyed are randomly assigned to Form X and the other half to Form Y.

    **Form X:** " Are we spending too much, too little, or about the right amount on *welfare*?”

    **Form Y:** " Are we spending too much, too little, or about the right amount on *assistance to the poor*?”

    The two questions ask essentially the same thing, but in different words. One year, of those given Form X, **23%** answered too little; of those given Form Y, **63%** answered too little.

    Research Question: "Does the wording of the question have an effect on the responses of American adults?"

    a. One variable of interest is whether the subjects in the study heard Form X or Form Y. Is this variable the predictor or the response variable? `Predictor variable, this Form is what defines the two groups to be compared`

    b. The other variable of interest is whether the subjects answered "too little", or not. Is this variable the predictor or the response variable? `Response variable, this is the outcome being compared`

    c. Describe the scope-of-inference for this study.

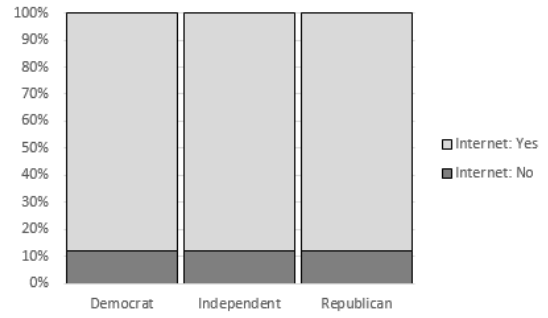
    `The scope-of-inference would be American adults. This is the group to whom the survey was given to.`

    d. Assume that the p-value from the appropriate analysis is 0.19. Write a conclusion in the context of the problem.

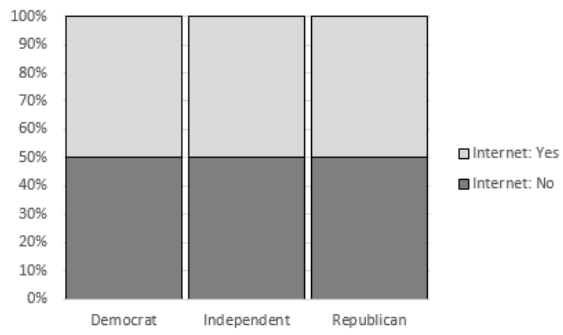
    `There is not enough statistical evidence to say the wording of the question had an effect on the responses of American adults (p-value = 0.19).`

Suppose these plots were used to investigate whether an association exists between one's political affiliation and whether they use the internet.

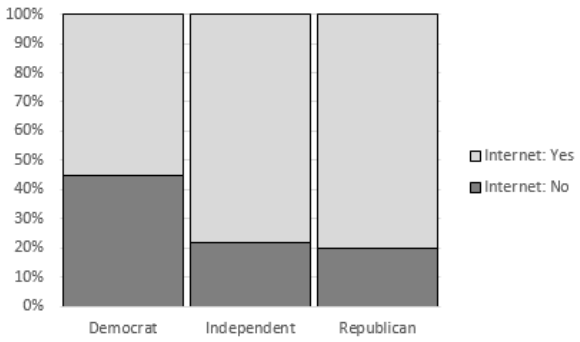
Plot I




Plot II




Plot III




21. Which mosaic plot (or plots) indicate that there is <u>no association</u> between one's political affiliation and whether they use the internet?

    a. Plot I only
    b. Plot II only
    c. Plots I and II
    d. Plot III only

22. The data shown in which plot would yield the smallest p-value?

    a. Plot I
    b. Plot II
    c. Plot III

23. A study published in the *Journal of the American Medical Association* investigated whether people taking statins (which are drugs to lower cholesterol) were more likely to have musculoskeletal problems. The data from the study were very similar to the following results:

| Group | Musculoskeletal Problems? | | *Totals* |
|---|---|---|---|
| | Yes | No | |
| Statin User | 6,061 | 906 | 6,967 |
| Not Statin User | 5,922 | 1,045 | 6,967 |
| *Totals* | 11,983 | 1,951 | 13,934 |

Compute the following quantities.

a. Find the risk of having musculoskeletal problems for the statin users.

$$\frac{6061}{6967} = 0.8699$$

b. Find the risk of having musculoskeletal problems for those not using statins.

$$\frac{5922}{6967} = 0.85$$

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

$$\frac{0.8699}{0.85} = 1.02$$

d. Find the odds of having musculoskeletal problems for the statin users.

$$6061 : 906$$

$$\texttt{Expressed as a ratio: } \frac{6061}{906} = 6.6898$$

e. Find the odds of having musculoskeletal problems for those not using statins.

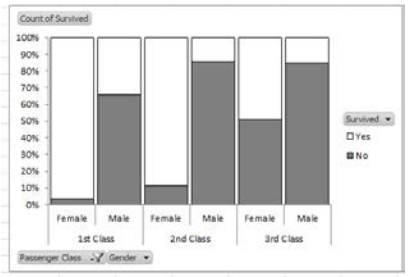
$$5922 : 1045$$

$$\texttt{Expressed as a ratio: } \frac{5922}{1045} = 5.6669$$

f. Find the odds ratio.

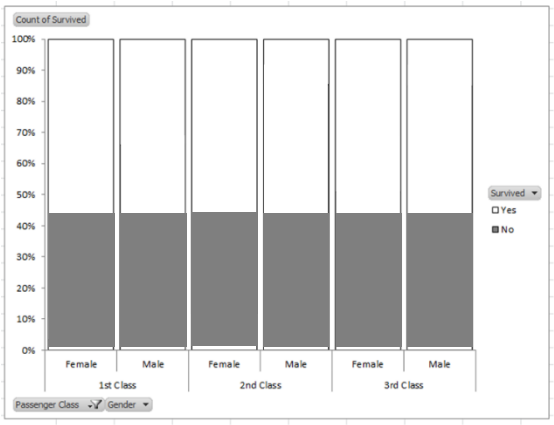
$$\frac{Odds\ of\ having\ musculosketal\ problems\ for\ statin\ users}{Odds\ of\ having\ musculosketal\ problems\ for\ those\ not\ using\ statin} = \frac{6.6898}{5.669} = 1.18$$

Consider the following summary table (i.e. PivotTable). 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.




24. 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.

Pattern must be straight across




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

Pattern different for Gender, but same for Passenger Class