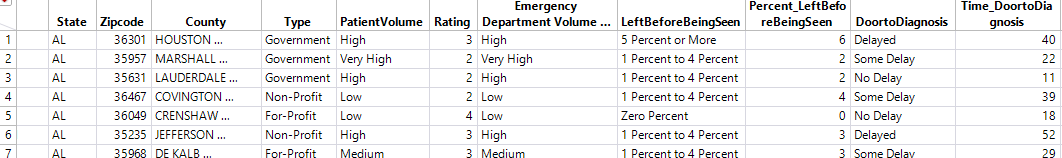
STAT 210: Exam #2 – Take-home Portion Name: Kevin Kemmerer  
Points: 25  
Fall 2018

|  |  |
| --- | --- |
| For this take-home portion of Exam #2, we will consider the Hospital Compare dataset. This dataset is published by Medicare.gov and provides a wide variety of outcomes regarding the quality of care provided at hospital across teh United States.  <https://www.medicare.gov/hospitalcompare/search.html>? |  |

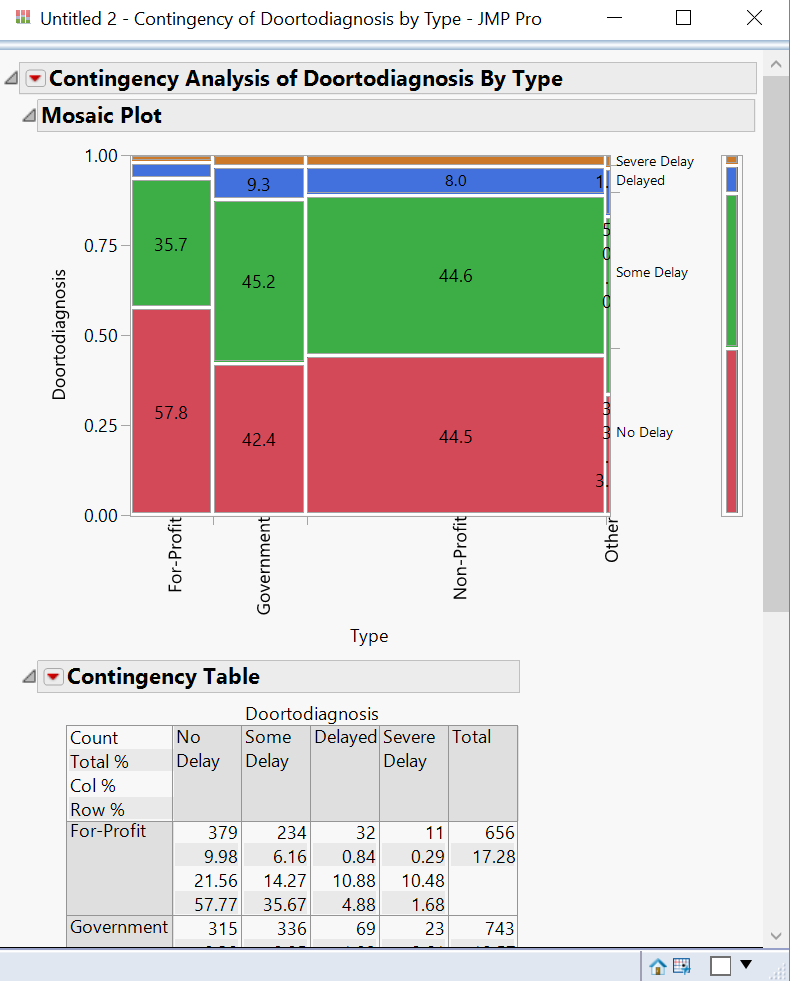
This dataset is provided on our course website. A snipit of the data is provided here.



The primary response of interest here is the average amount of time it takes from walking into an emergency room until a diagnosis is provided to the patient. This variable, DoortoDiagnosis, is described here.

|  |  |
| --- | --- |
| Labels for  DoortoDiagnosis | Amount of Time from Walking in Door to Diagnosis |
| No Delay | *Less than 20 minutes* |
| Some Delay | *20 – 39 minutes* |
| Delayed | *40 – 59 minutes* |
| Severe Delay | *60 minutes or more* |

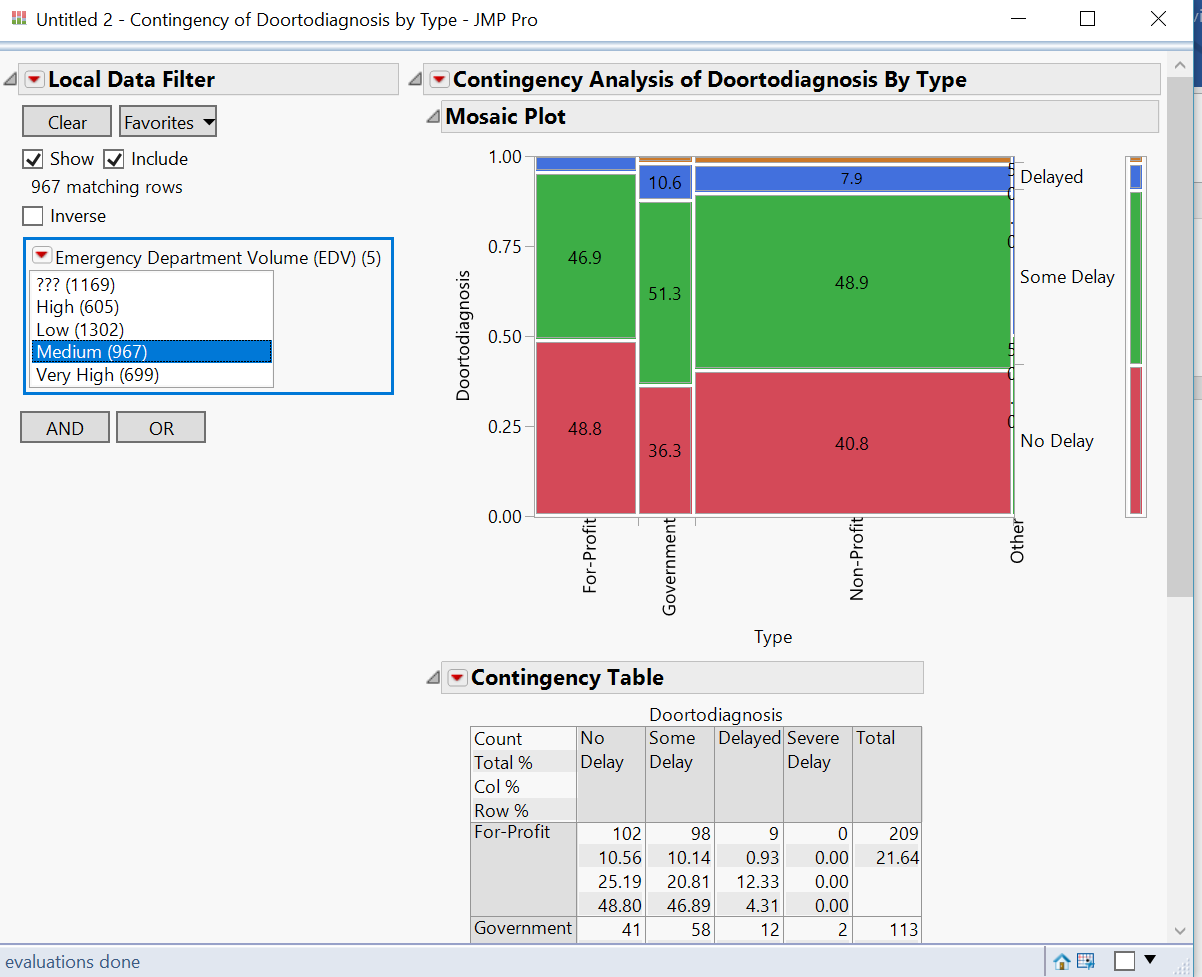
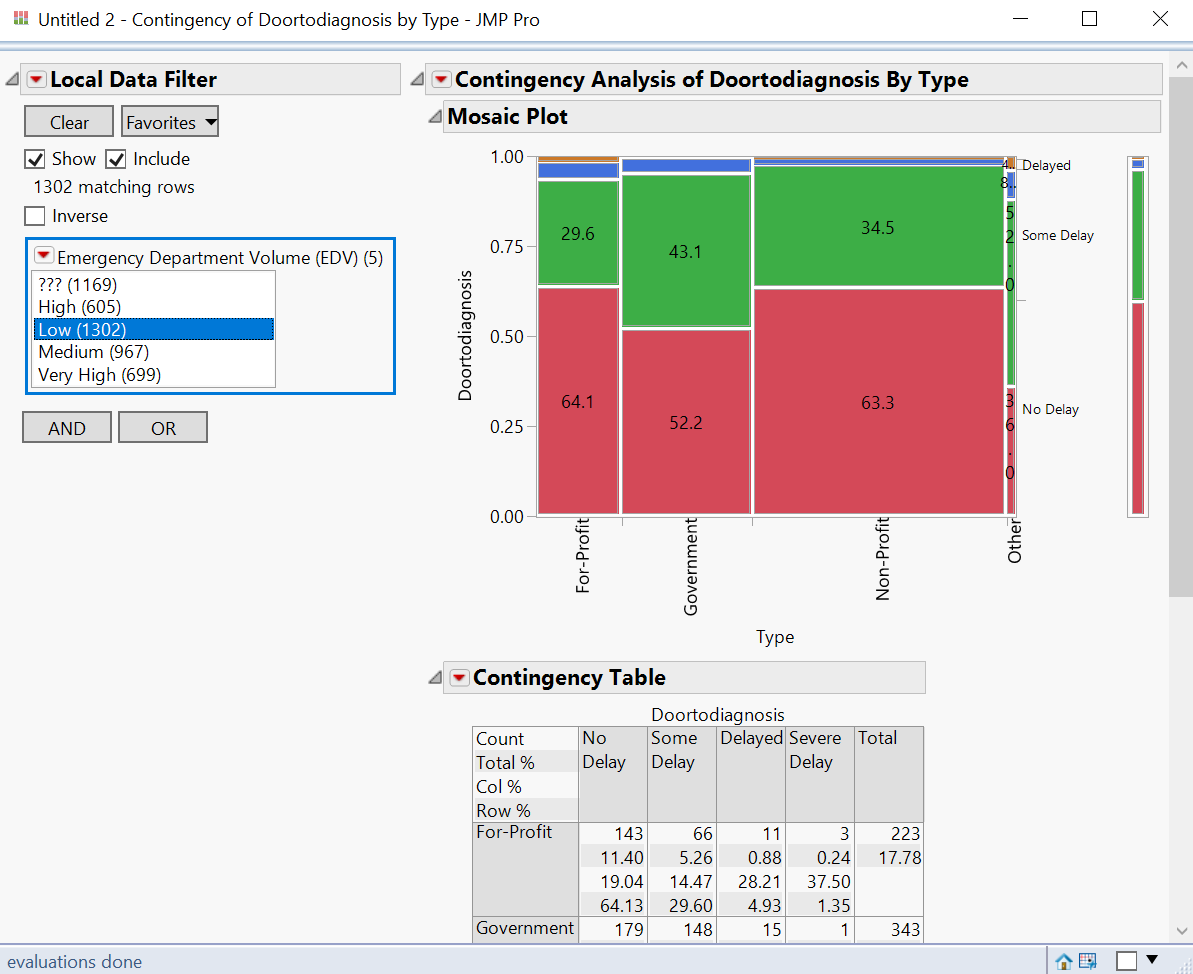
1. Create a mosaic plot to show the association Hospital Type (For-Profit, Non-Profit, Government) and DoortoDiagnosis. DoortoDiagnosis is the primary outcomes, i.e. the response variable here. Fix the value ordering for DoortoDaignosis so the outcomes are easy to read.

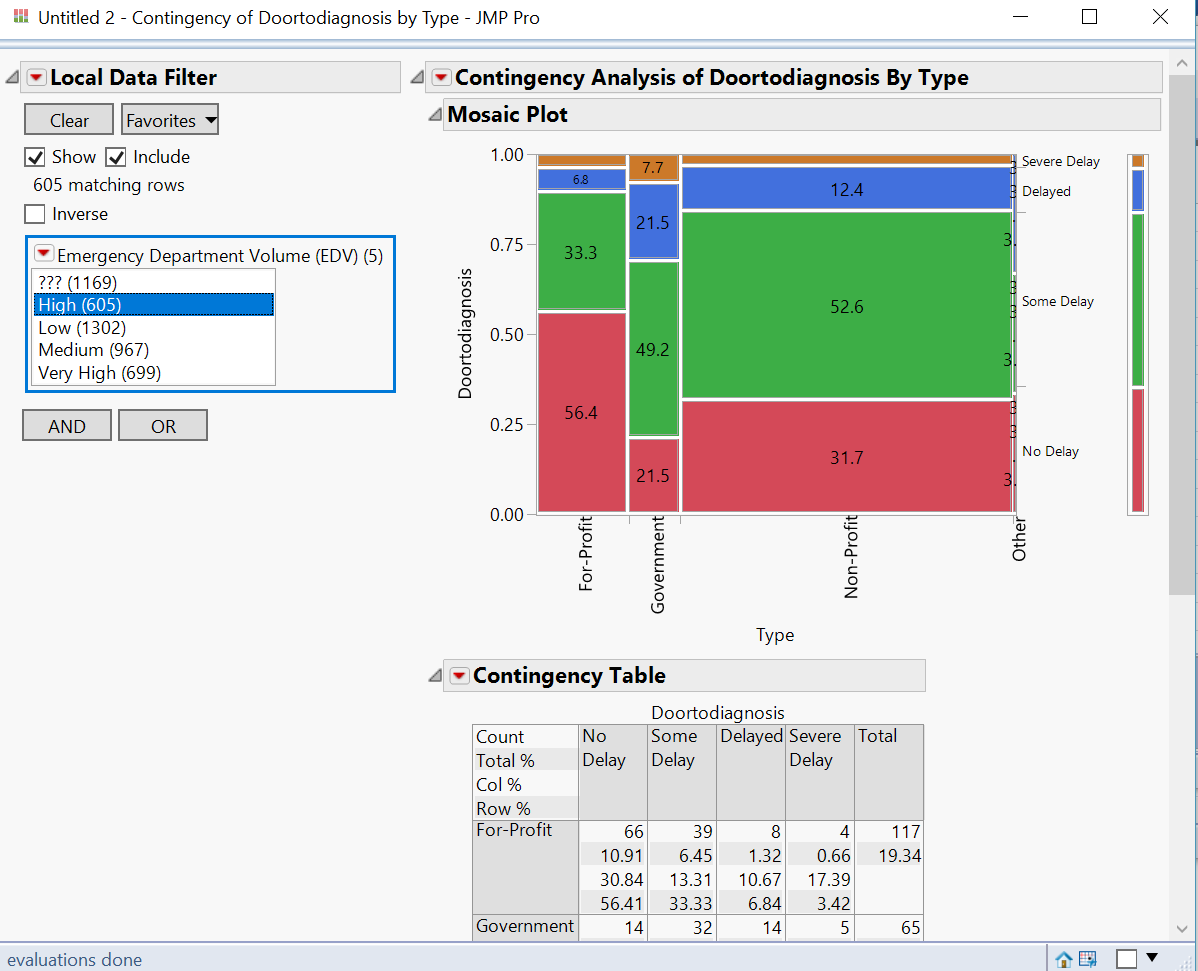
***<Screenshot your mosaic plot and paste it here>***

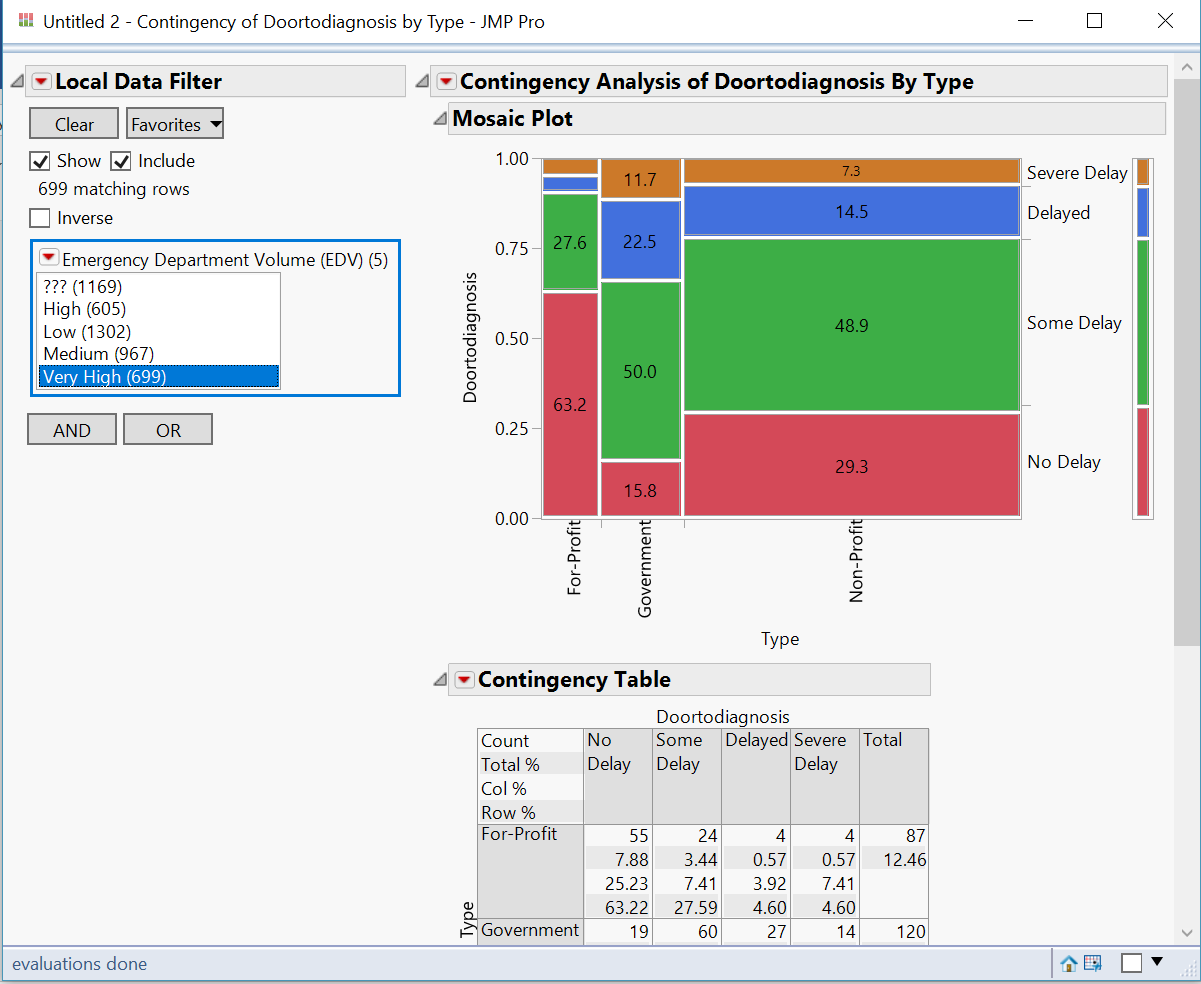
Which hospital type has the largest proportion of No Delays? (2 pts)

For-Profit type hospitals have the largest portion of No Delays.

1. Describe the association between Emergency Room Volume and DoortoDiagnosis.

***<Screenshot your mosaic plot and paste it here>***



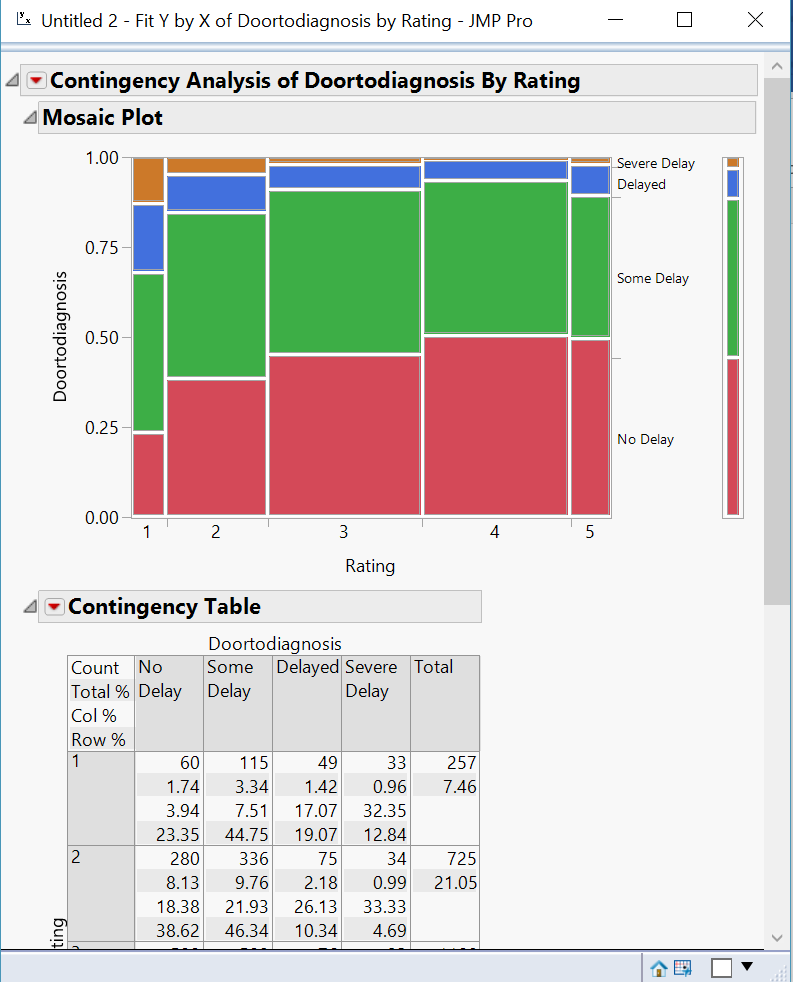


Do the patterns that you see in this plot make sense? Briefly discuss. (2 pts)

Yes, for the most part as the EDV increases the delays continue to get longer.

1. Create a mosaic plot between Rating and DoortoDiagnosis.

|  |  |
| --- | --- |
| You will need to change the Modeling Type on Rating to Nominal (instead of Continuous) in order to create the appropriate graphic. |  |

***<Screenshot your mosaic plot and paste it here>***

* 1. Conduct a quick Google search, what is considered a good hospital rating? What is a poor rating? (1 pt)

Only hospitals that have at least 3 measures within at least 3 measure groups or categories are eligible for a hospital rating. 1 star is the worst rating, and 5 star is the best rating for hospitals in this rating scale.

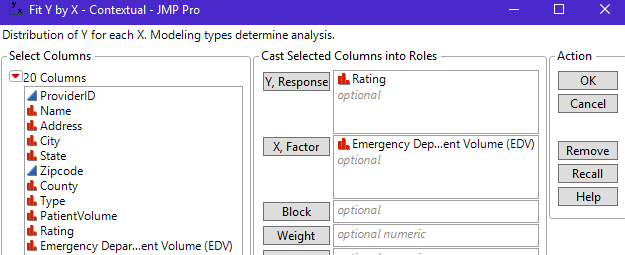
* 1. Does it appear that DoortoDiagnosis is one of the factors that contributes to a hospital’s overall rating? Briefly discuss. (2 pts)

Yes, from the data provided it does look like the timeliness of care contributes to a hospitals overall rating. Also, online it shows that timeliness of care is taken into consideration when rating a hospital.

* 1. Do the patterns that you see in this plot make sense? Briefly discuss. (2 pts)

Yes, as the rating increases there is a slight decrease in delays across the board.

1. In JMP, select Fit Y by X and specify the following. (Again, make sure Rating is a nominal data type).



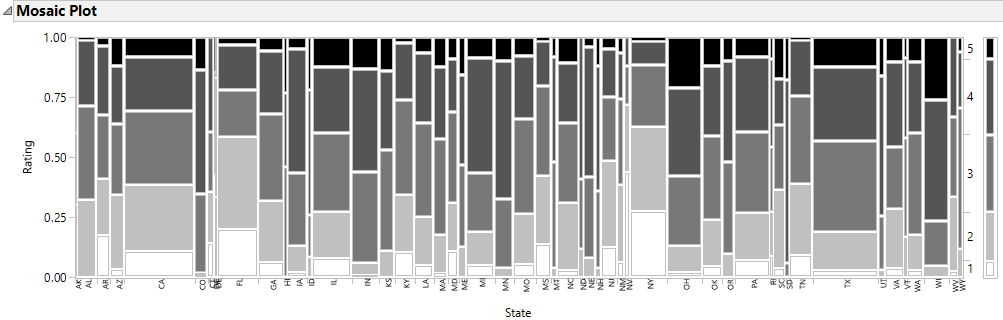
* 1. Determine whether or not the data supports the following statement. (2 pts)  
      “As emergency room volume increases, the likelihood of receiving a Rating Score of 1 increases

True, the data does support this statement.

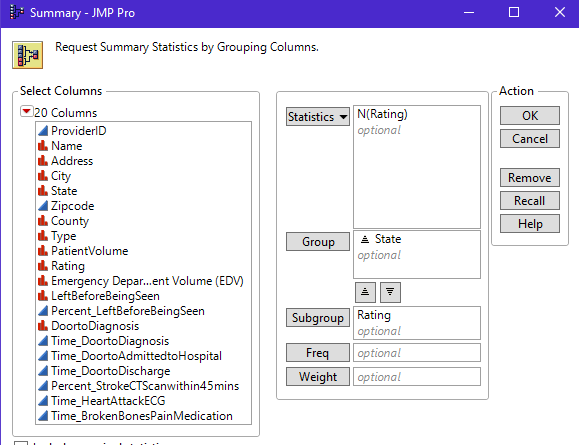
* 1. Determine whether or not the data supports the following statement. (2 pts)  
     “As emergency room volume increases, the likelihood of receiving a Rating Score of 5 decreases.”

False, it seems that the hospitals with a rating score of 5 maintain that score no matter what the emergency room volume is.

Let us next consider a comparison between the Hospital Rating Scores across States. From the following graphic, it appears that WI has the largest proportion of hospitals with a 4 or 5 rating score, MN has as a large proportion of hospitals with a rating score of 4 or 5.

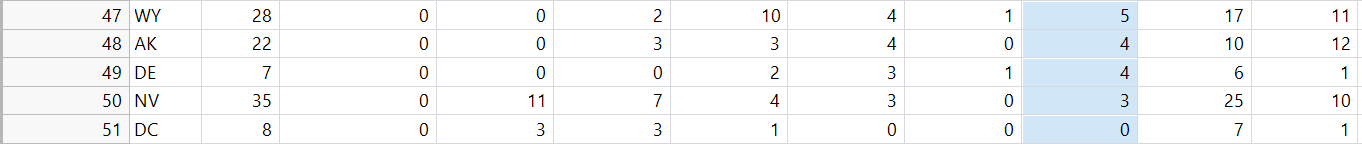
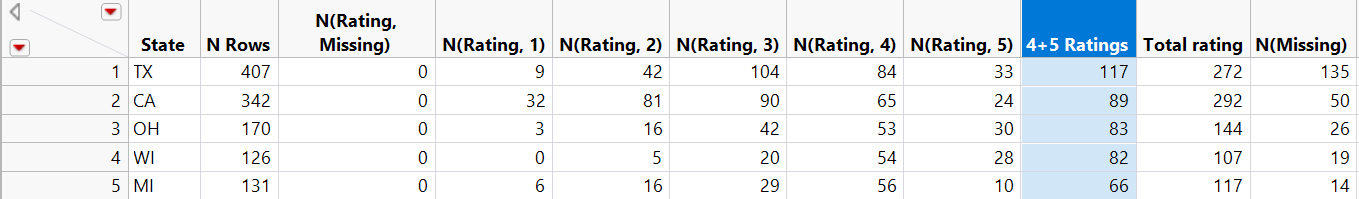


In an effort to rank states based on hospital rankings, let us consider a table summary in JMP. Select Tables > Summary. Specify the following setup and click OK.



1. Fill-in the following tables. (3 pts)

* Top 5 States represent the states with the highest proportion of hospitals that have a rating score of 4 or 5
* Bottom 5 States represent the states with the lowest proportion of hospitals that have a rating score of 4 or 5
* When ranking, the total number of ratings should be used -- not the number of rows (some hospitals do not have a rating score)
* You can treat DC as “state” for the purposes of ranking

***<Take a couple screenshots showing your top 5 & bottom 5 rankings and paste it here>***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Top 5 States | |  | Bottom 5 States | |
| 1. 117 | TX |  | 47. 5 | WY |
| 1. 89 | CA |  | 48. 4 | AK |
| 1. 83 | OH |  | 49. 4 | DE |
| 1. 82 | WI |  | 50. 3 | NV |
| 1. 66 | MI |  | 51. 0 | DC |

More and more people are interested in these types of rankings as quality of care becomes increasingly important as healthcare expenditures continue to rise. On way a state could artificially increase their ranking is to have hospitals with poor performance not submit the necessary data so that a rating score could not be determined for that hospital.

Comment: There are a variety of reasons a hospital may not get a rating score and I am not accusing any state of falsely withholding information in order to increase rating scores. This is a hypothetical situation.

1. Answer the following (1 pt each)  
   1. What State has the highest proportion of missingness for rating score?

TX has the highest proportion of “missingness” for rating score.

* 1. Is this state ranked in the Top 5?

This state is ranked the best.. out of the top 5 states

* 1. If the statement regarding with-holding information is true, should this state be listed in the Top 5? Briefly discuss.

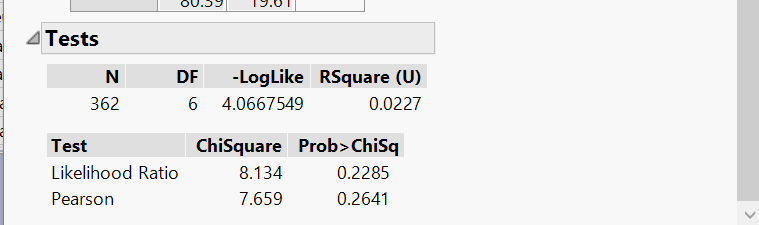
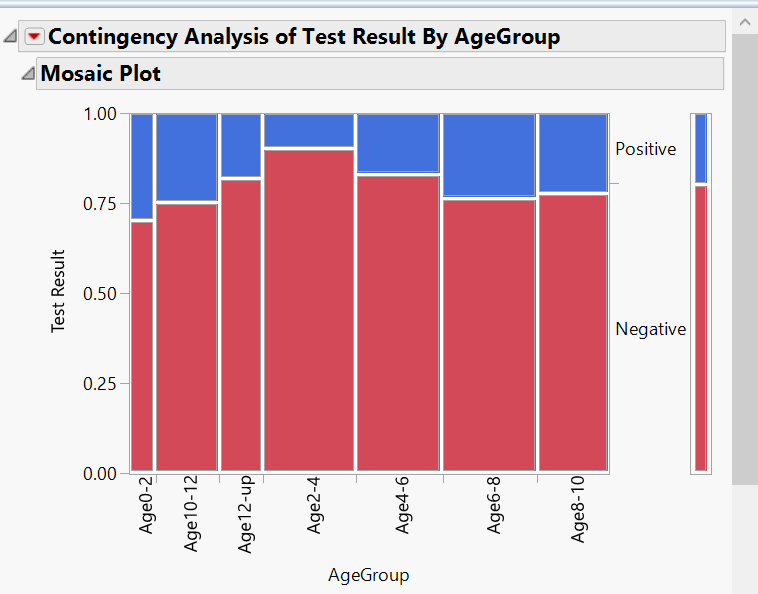
No, this state should not be listed in the top 5 because this state may have withheld information to boost the total rating scores.

|  |  |
| --- | --- |
| For the last two problems, consider the Dogs\_LymeDisease dataset provided on the course website. This data is from an animal shelter and provides information on lyme’s disease testing outcomes on dogs. | Image result for lyme disease dogs |

1. Conduct a Chi-Square Test of Independence for the following hypothesis.

*Note*: The Age Group variable will suffice for Age of Dog.

***<Screenshot your JMP output (including the p-value) and paste it here>***

******

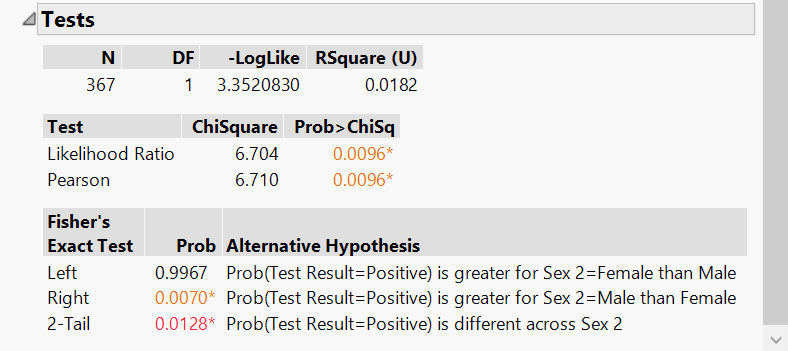
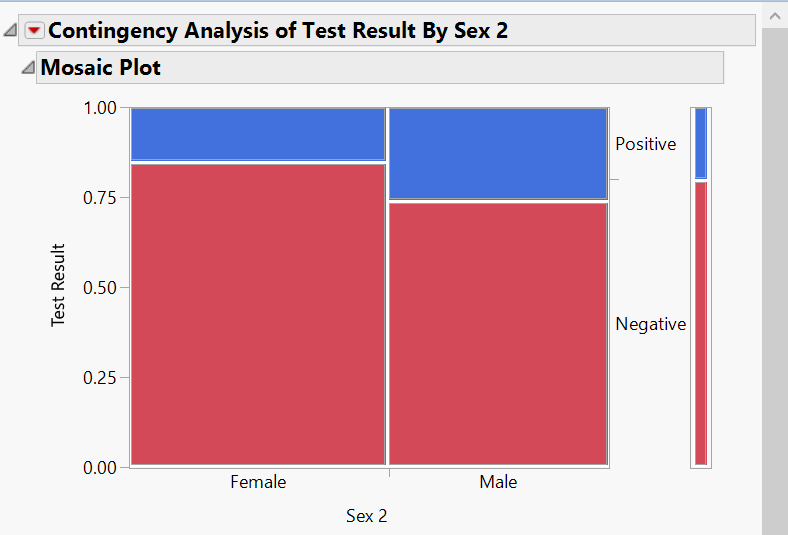
Write a conclusion for the above test of hypothesis in laymen’s language. (3 pts)

In conclusion, in this test where the alternative hypothesis is the likelihood of a positive test result being related to the age of the dog we have to accept the null hypothesis. The probability of having a dog test positive is p value (.2285) for the test to be statistically significant it must have a p value of (.05) 5% or less. The age of the dog has no statistical significance when compared to the likelihood of a positive test result.

1. Conduct a Chi-Square Test of Independence for the following hypothesis.

*Note*: Use Cols > Recode to recode the Sex variable as described below before running this test. This new variable should only have two categories for sex.

|  |  |
| --- | --- |
| Original Sex Variable | Recoded Sex Variable |
| Female | Female |
| Spay |
| Male | Male |
| Neuter |

***<Screenshot your JMP output (including the p-value) and paste it here>***

Write a conclusion for the above test of hypothesis in laymen’s language. (3 pts)

In conclusion, in this test where the alternative hypothesis is the likelihood of a positive test result being related to the sex of the dog, we reject the null hypothesis. The probability of having a dog test positive based on sex p value (.0096) for a hypothesis to be deemed statistically significant it must have a p value < (.05) or 5% which our test does. The sex of the dog does have statistical evidence when compared to the likelihood of the dog having a positive test result.