

**Data Analysis – EIT – Data Science**  
**UPM**  
**Assignment 2.2**  
**Dina Mohamed**

- 1- Y : Elks crossing road  
 X: Car or Truck  
 Z: low or high traffic

➤ The odds ratio for vehicle and action by traffic can be seen below:

<i>odds ratios for vehicle and Action by Traffic</i>	
<i>High</i>	<i>Low</i>
<i>5.2868421</i>	<i>0.9595142</i>

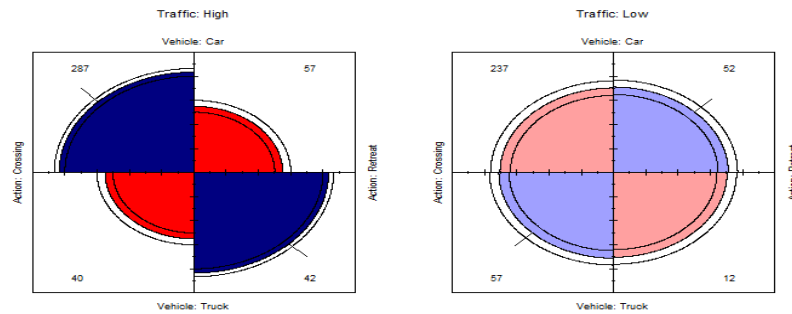
Interpretation: Taking into account the traffic level, we can say that the odds ratio for an elk to cross the road in front of cars is 5.2 times higher than in front of trucks

Confidence intervals can be seen as below:

	<i>odds</i>	<i>2.5 %</i>	<i>97.5 %</i>
<i>Car:Truck/Crossing:Retreat/High</i>	<i>5.2868421</i>	<i>3.1495667</i>	<i>8.874459</i>
<i>Car:Truck/Crossing:Retreat/Low</i>	<i>0.9595142</i>	<i>0.4807475</i>	<i>1.915075</i>

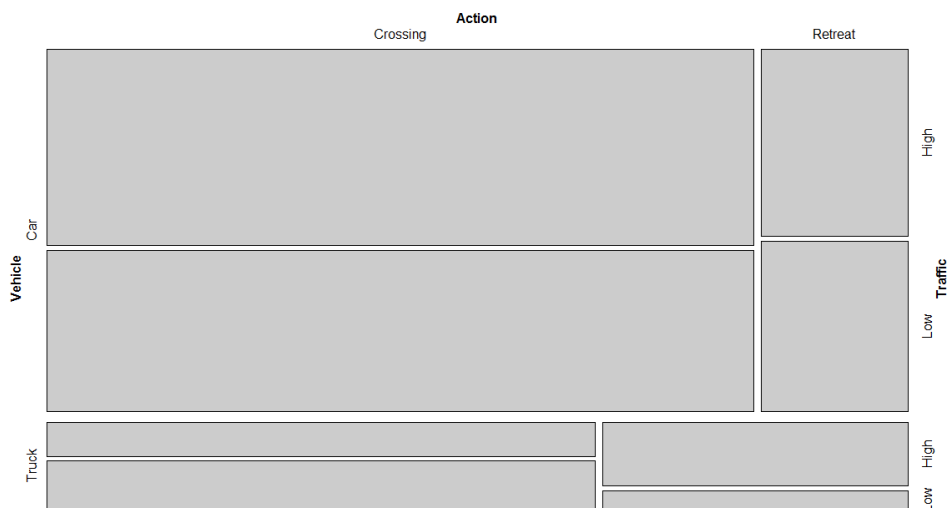
Interpretation: This means that the odds of crossing elks for cars vehicles are higher than trucks

- Plotting the fourfold , as in the below picture, we can conclude:
- For the first plot, there is a strongly positive association for high traffic level, the rings don't overlap and the non-principal diagonal sectors have less area than the principal diagonal ones (odds ratio is 5.2). We can reject the hypothesis of independence between the variables.
  - For the second plot, there is a weak negative association for the low traffic levels. We cannot reject the hypothesis of independence between the variables (Variables look independent for low traffic , odds ratio is 0.95)



In the below graph we can see the Mosaic plot, where we can conclude the below:

- The number of elks crossing in front of a car is much more higher than those crossing in front of a truck (this can be seen as the width of the crossing is bigger than the width of the retreat)
- The number of elks crossing in front of a car doesn't vary much from high or low traffic
- The number of elks crossing in front of a car DO vary from high or low traffic
- The number of elks crossing in front of a truck is less in high traffic than in low traffic

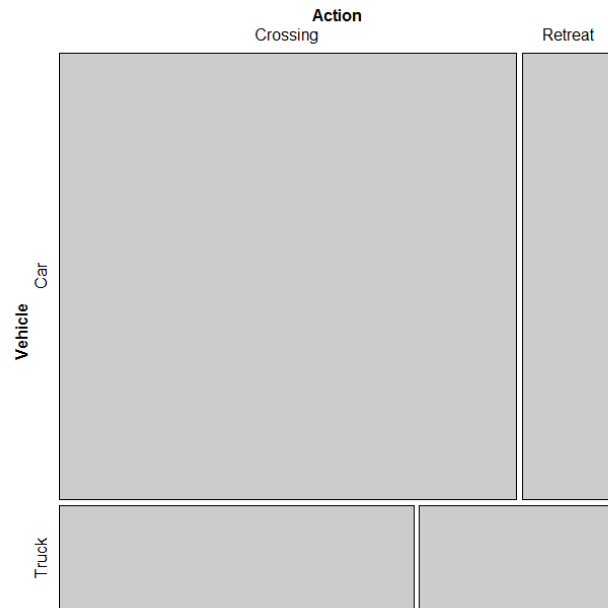


- 2- Calculating the oddsratio taking into account the elks crossing vs retreating without taking into account the traffic level, results can be seen below:

*odds ratios for vehicle and Action*

*[1] 2.676251*

Interpretation: The odds of elks crossing the roads in front of cars are higher than in front of cars, we can also see this if we plot the mosaic against the marginal table as below:



As in the above image, we can see that the width of the bin of of elks crossing in front of cars is much bigger than the width of the bin of elks crossing roads in front of trucks.

We can conclude that using the marginal or partial tables doesn't differ, both tables give the same conclusion which is: elks crossing roads in front of cars is higher than in front of trucks.

- 3- To test the hypothesis of homogeneous association (or homogeneity of odds ratio) for elks crossing the road we will use the "woolf\_test" function. Result can be seen as below:

*woolf-test on Homogeneity of Odds Ratios (no 3-way assoc.)*

*data: ElkCrossing.table.partial*

*X-squared = 14.999, df = 1, p-value = 0.0001076*

Since P-value is very small, with confidence of 5%, then we can reject the homogeneous association which means that the effect of elks crossing the roads in front of cars or trucks is not the same at each level of high or low traffic. We can reject the Homogeneity of the Odds Ratio through the two categories of traffic level.

- 4- To check if X and Y are conditionally independent on Z we use the Mantel-Haenszel chi-square function as illustrated below

*Mantel-Haenszel chi-squared test with continuity correction*

*Data: ElkCrossing.table.partial*

*Mantel-Haenszel X-squared = 24.39, df = 1, p-value = 7.868e-07*

*Alternative hypothesis: true common odds ratio is not equal to 1*

*95 percent confidence interval:*

*1.801123 3.924165*

*Sample estimates:*

*Common odds ratio*

*2.658553*

We can see that the p-value is very small, so we can reject the null hypothesis which means that there is a conditional dependence between the vehicle type and elks crossing roads for each type of traffic, that means that true common odds ratio is not equal to 1.