

## Analysis of the Accident Dataset

We have the dataset of the road accidents in Great Britain in 2010. There were 32 variables collected. We will analyze the difference in the mean number of accidents by months and show the histogram of hour of accidents.

We do the hypothesis test. There is **NO** enough evidence to say that the mean number of accidents by day for each month are indifferent. (F-stat=11.29, p-value<0.0001) The figure 1 shows the estimated mean number of accidents by month along with 95% confidence interval. **November** seems have the highest mean (484.8 and IC: (457.7-511.9)) and we need deploy more resource to deal with large number of accidents in Nov. Additionally, the figure 2 shows that the accidents tend to occur at 7:30-8:30 and 15:30-17:30 It is not a surprise because people are going to work or coming back home at this time. It is the rush hour.

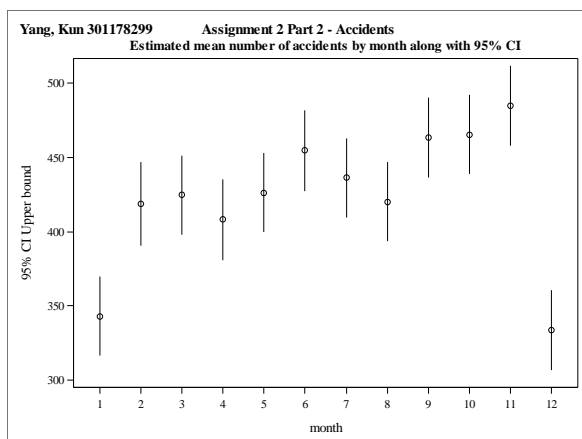


Figure1: It shows estimated mean number of accidents by month along with 95% CI

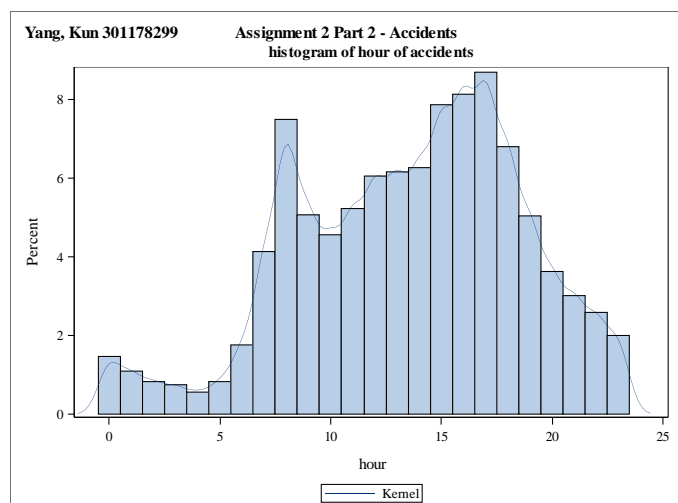


Figure2: a histogram of the minute of the hour for accidents