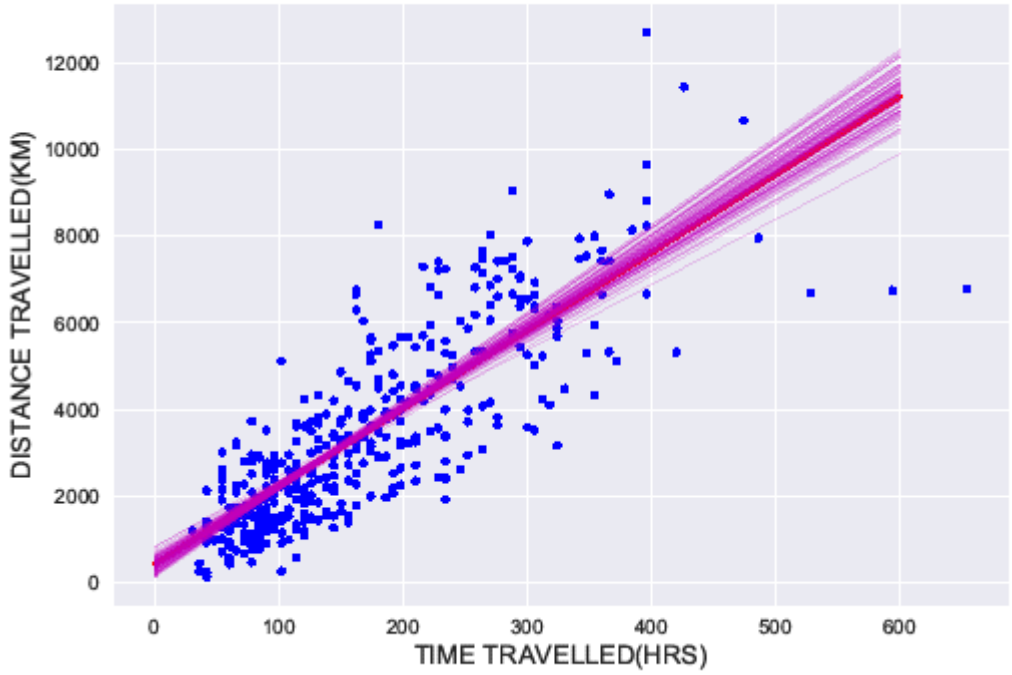
# Part 5: Hurricanes and Emergencies -- Inferential Statistics

1) The first question we would like to ask is how random the average speed of Atlantic hurricanes is? The column 'AVERAGE SPEED(KMPH)'is obtained by taking the relation of the 'DISTANCE TRAVELLED(KM)' value to the 'TIME TRAVELLED(HRS)' value, so that the values in the 'AVERAGE SPEED(KMPH)' column are determined by the values of the two other columns. But how independent is the time travelled by a hurricane from the distance travelled? The mean value of the average speed of Atlantic hurricanes is 21.132572425720934 km per hours, while the standard deviation is 8.538384619924566 km per hour and is quite large compared to the mean.

The distance and time travelled columns are strongly correlated; the Pearson coefficient is 0.813291065794.

We then make the scatter plot of the distance travelled vs the time travelled and make the linear fit. We see that the slope of the fit, that is in fact the fitted velocity of the hurricane motion, is 17.9969045062 km per hour and is slightly less than the mean value of the average speed 21.132 km per hour.




It is instructive also look at the 95% confidence interval for the slope using pairs bootstrap method and visualize the variability of the slope. To do this, we use the auxiliary function that generates the replicates of the slope and intercept. The confidence interval for the slope (fitted velocity of a hurricane) is [16.01, 20.29]

km per hour. We see that the mean value of the average speed being equal to 21.13 km per hour is slightly higher than the upper bound of the confidence interval.

2) We can then look at the Pearson coefficients between some other columns describing independent variables.

a) We see that the maximum effective strength of a hurricane is strongly correlated with the average strength with respect to maximal; the Pearson coefficient is -0.657 and negative. This means that the larger the maximum strength achieved by the hurricane, the smaller amount of time it spent being that strong. This is in complete agreement with the visualization data presented in the data storytelling section.

b) It is completely natural that the maximum effective strength is strongly positively correlated with the average wind strength inside a hurricane (Pearson coefficient is 0.858). This is because the strength of a hurricane at any moment of time is primarily based on the speed of the wind.

c) It may seem first a bit surprising that the first latitude is rather strongly correlated with the first longitude when a hurricane first becomes dangerous (Pearson coefficient is -0.485). This correlation may be explained by the presence of land (islands and continents) that affects the place of formation of hurricanes.

d) Finally, it's worth looking at the correlation between the maximum effective strength and first latitude and longitude when a hurricane first became dangerous. The correlation between the maximum strength and longitude is rather weak (Pearson coefficient is 0.268), while the correlation involving latitude is stronger (Pearson coefficient is -0.324) and negative. The latter correlation can be explained by the fact that the strongest hurricanes form closer to the equator and gain power during their motion above the ocean.

3) Finally, we will consider the correlation between the dependent binary variable of the 'LED TO EMERGENCIES' column and some of the independent variables.

a) The binary dependent variable is strongly correlated with the maximum effective strength variable that is also discrete (Pearson coefficient is 0.517), as one could see from the data storytelling part. The correlation of emergencies variable and the average strength with respect to maximal is somewhat weaker (Pearson coefficient is -0.324)

b) From the other side we see that our dependent variable is not very strongly correlated with distance (Pearson coefficient is 0.268) and time (Pearson coefficient is 0.280) travelled by hurricanes. Let's formulate the null hypothesis that these variables are correlated and calculate the corresponding p-values using permutation method. They are equal to 1 meaning that both null-hypothesis are true.

c) Looking at the correlation between the dependent ‘emergencies’ variable and first latitude and longitude variables, we see that the target variable is weakly but correlated with the first latitude variable (Pearson coefficient is -0.178) and almost uncorrelated with the first longitude (Pearson coefficient is 0.022). Let's test two null-hypothesis: the presence of correlation in the case of latitude and the lack of correlation in the case of longitude. The computations below show that in both cases we obtain large p-values, meaning that both null-hypothesis should be kept. We thus come to a remarkable conclusion that the value of the first latitude affects the probability of emergency declarations, but the value of the first longitude does not.