Good afternoon all. My name is Andrew Estes and my thesis is titled a Poisson Analysis of Formula 1 Pit Stops. Basically, I attempted to predict when a pit stop would be made during a Formula 1 race.

To start with, we need to look at data. The type of data we have leads us to different models. My goal was to predict whether a pit stop would be made or not. This clearly indicates towards a Poisson Distribution in the probability that an event – pit stop in this case – occurs.

Now, let’s define Poisson. In probability theory and statistics, the Poisson distribution is a discrete probability distribution – gotcha! I won’t be reading directly from the slides.

To keep it simple, a Poisson distribution involves data that is countable. Height based upon gender is a good example, there are 120 thousand women with the height of 160cm. Another type would be number of pit stops during a race.

Okay, now we should have an expert-level understanding of statistics at this point. You’re welcome and tips can be left in the jar on the way out. With our statistical expertise accounted for, we now need to learn about Formula 1, aka F1. F1 cars are the fastest road-racing cars in the world. In 2020, a driver Lewis Hamilton averaged 164 mph through a lap and back in 2005, a driver hit a top speed of 232 mph. However, it isn’t a drag race. There are many corners that require significant braking – slowing to 60mph before re-accelerating up to 200mph. The impact of the braking is often measured between 4-6g force. To put that in perspective, Apollo 16 – the last mission to have men on the moon – re-entered the earth’s atmosphere at 7g forces. What we see here is the starting grid. There are 20 cars made up of 10 separate teams.

Those drivers are all competing with each other for placement. The higher you place, the more points you get. And a cumulative point total leads to a driver’s final placement at the season’s end. The team’s final placement depends on the cumulative point total from both drivers. As you can see in 2022, McLaren lost to Alpine even though McClaren’s top driver was better than Alpine’s top driver.

There are over 20 races per season and normally five of the six hospitable continents are utilized. Within each continent, many different countries have provided a course.

Within each country, there are multiple tracks that can be used. You can see the US has the highest variety of tracks and is the fourth most frequently visited country.

Of the 27 countries visited in the 5-year period studied, Italy was, by far, the most frequently visited country. This is likely due to the heavy racing that took place in 2020, the year of Covid.

With the wide variety of continents and countries, there is a wide variety in climate. You can see the basic weather numbers here…

In addition to the global-trotting schedule, each course has a unique layout. You can see the France and Chinese course layout here. The colors correspond to the gear shift of the driver who had the fastest lap in the respective race. China is obviously much curvier and requires more braking than France. Looking at the color-pattern, the fastest driver in China had to go down to 2nd gear one time and 3rd gear three times. Meanwhile in France, the lowest gear was 3rd gear which was hit three times. Similarly, the highest gear in France was hit 3 times, in China, it was only hit once.

Not only does the layout change, but there are two types of tracks. There is a dedicated race track and there is a street track. Here is a snapshot of the Vegas race that will occur in November this year. It’s right on the strip. And as you can imagine, having a race on a regular road has it’s own challenges that dedicated tracks don’t provide.

Let’s look at Monaco for example. This is a steep decline with a tight turn.

Austin’s “tight” corner isn’t as angled for the turn, nor does it have the elevation change to consider.

My analysis was specifically on the pit stop aspect of the race. In this picture, you’ll see the pit stop lane on the left side – and can even make out a car leaving the pit stop area. All pit stop lanes are at the start/finish line. And while it looks clear-cut, it can be pandemonium.

Here is a photo of two cars leaving at once along with a third car getting the tires changed.

It’s quite crowded. In fact, there are 20 people for each car with specific roles when it comes to the pit stop.

Here is a real-life picture of a pit stop with the various crew members huddled in.

Now that we have an idea of what a pit crew responsibility is, we can actually look at a video of a pitstop. Man, it would be great to be able to change the tires on my car in 2.1 seconds instead of making a 4 hour long appointment. This is completely unrelated to the analysis but if it took 2 seconds to change a vehicle’s tires, the entire population of Kirksville could have their tires changed in less than 10 hours.

Back to the video, I took a snippet 1 second apart. Excluding the slightly separate camera angle, can you tell me the difference here? It's the stripe on the tire. That indicates that a different compound is used.

In fact, there are 6 different types of hard tires, excluding intermediate tires and wet tires.

The type of tire used, and the decision to make a pit stop, is determined by a group of personnel including the race strategist, chief engineer, other crew members, and to an extent the driver. The staff has a massive influx of data coming from the various sensors on the car to help measure tire degradation and other information. And they use this information to make the decision to pit or stay out.

Other than the tires and timing, a consideration a strategists uses to make a pit stop decision involves flags. A yellow flag forces drivers to drive no faster than 80 kmh, around 50mph. If drivers on the course are only going 50mph, rather than 150mph, then the time loss suffered by making a pit stop is significantly reduced. So a 20 second loss can be transformed into a 15 second loss.

Now, let’s take a look at the F1 race in Austin. Any guesses where the start-point is?

Any questions for me?

The first step in any project is getting the data. The data for the F1 information can only be accessed via Python. Thankfully I took Dr. Shanshan’s class as it helped me understand how to get the data. Here is a snippet of the over 500 dataframes for the initial analysis. You’ll notice a wide variety of dimensions. Even the same race a year apart has different dimensions.

For the laps dataframe, there were 27 columns and rows ranging from 60 to over 1000 for each race.

The results dataframe had 16 rows and 20 columns for each race.

We combined those two dataframes by Driver, Grid Position, Position, and Points to create a “final” dataframe.

There are two other dataframes that we considered but did not use. The first was race control messages.

The second was weather.

Here are some snapshots of various transformations.

Now that the data has been wrangled, we did some visualization in Shiny. First is the nicely colored one showing the frequency of pit stops by team at a specific course over the 5-year period.

Here is a histogram of tire life when pit stop was made. Pretty normal to right-skewed distribution. And here is the lap number when a pit stop is made – more right skewed.

We then looked at the mean-variance for Saudi Arabia 2018. And then again by tire grouping. The mean-variance for the full dataset was even worse (xxx and xxx)

Then we created a pit stop distribution dataframe to look at the frequency of pit stops, tire life, and when they were made.

Here is the visualization, both actual and normalized

Time for Poisson analysis

Consul’s Poisson

Famoyes Poisson

Negative Binomial

Alpha calculation

Pretend everything’s all right

Predicted vs Actual analysis

Random Forest

Other models

Bayesian Process