Formula 1 Statistics Research Project

Christopher Schretzmann

Final Research Project

Probability & Applied Statistics

Project Topic and Background

For my project, I decided to focus on Formula 1 racing since there is a lot of data throughout the years. Formula 1 is not only a test of athletic performance but also engineering. Unlike other types of racing, the engineering aspect is part of the formula as well. Each team gets two drivers on the grid for each season and for each race, depending on the position, the drivers each gain points. Each location of the race consists of a whole race weekend. There are set practice times each team gets to run. Since engineering is part of the formula, each team is limited to a certain budget and practice time for each season. Some teams are very large with up to 1200 employees just for two drivers. That is one thing that drew me to the sport because there are many variables and outcomes. Formula 1 has been around since the ’50s and continues today. For the purpose of this research project, I decided to cover the years 2015 to 2022 since the differences in engineering breakthroughs for the industry vastly changed outcomes. It must be noted that 2021 is the year where they started the spending budget rules limiting each team to 145 million per year. The athletes must have the endurance of a marathon runner and the strength of a bodybuilder to withstand the forces created by the futuristic racing machines. They can lose up to 10 pounds per race since the conditions are so hard on their bodies. We are going to follow Max Verstappen, a driver who started racing Formula 1 in 2015 and is currently dominating the sport for the past few years. This time period is the “hybrid era” for Formula One where they are required to spec their engines to be hybrid engines. Over time they change their requirements in every aspect of the sport and the engineers must innovate within those boundaries.

Chapter 2

The sets for the 2023 season are as follows

Each Team consists of two racers and all of the racers are from all over the world. Most of the racers are European, but some are from South America. There is only one racer from the United States. Each year, the roster is different, so I used the 2023 year.

SET THEORY

European Drivers = {Max Verstappen, Carlos Sainz, Charles LeClerc, Lando Norris}
RedBull = {Max Verstappen, Sergio Perez}
Ferrari = {Carlos Sainz, Charles LeClerc}
McClaren = {Lando Norris, Oscar Piastri}

RedBull OR Ferrari = {Max Verstappen, Sergio Perez, Carlos Sainz, Charles LeClerc}

RedBull AND European Drivers = {Max Verstappen}

RedBull NOT European Drivers = {Sergio Perez}

Demorgan’s Law States

NOT Red Bull AND NOT Ferrari = NOT(RedBul or Ferrari)

Central Tendencies

Max Verstappen is a very consistent driver, even including his rookie year.

Max Verstappen Stats

FORMULA ONE STATS

YEAR	RANK	STARTS	WINS	POLES	TOP 5	TOP 10	PTS	AVG START	AVG FINISH
2015	12	19	0	0	2	10	49	--	--
2016	5	21	1	0	11	17	204	--	--
2017	6	20	2	0	12	13	168	--	--
2018	4	21	2	0	15	17	249	--	--
2019	3	21	3	2	17	19	278	--	--
2020	3	17	2	1	11	12	214	--	--
2021	1	21	10	8	19	20	413.5	--	--
2022	1	22	15	7	17	20	454	--	--

<u>The Average Position</u> for the end of the season for Max Included in the years covered is 4th place

$$(12+5+6+4+3+3+1+1) / 8 \text{ years} = 4.375$$

<u>The Variance</u> of his performance is

$$\frac{(12-4.375)^2+(5-4.375)^2+(6-4.375)^2+(4-4.375)^2+(3-4.375)^2+(3-4.375)^2+(1-4.375)^2+(1-4.375)^2}{8}$$

The Standard Deviation for his performance in these years is

$$\sqrt{\frac{(12-4.375)^2+(5-4.375)^2+(6-4.375)^2+(4-4.375)^2+(3-4.375)^2+(3-4.375)^2+(1-4.375)^2+(1-4.375)^2}{8}}$$

To find The Median of the values we have to sort in ascending order. Since there is an even number of years, the two values in the center are averaged

1,1,3,3,4,5,6,12

The Average of 3 and 4 is 7/2 or 4.5

There is no mode for this data because one number does not occur the most.

Probability

If there are 20 cars on the track in each given race, What is the probability that Max could win the race if there is an equal chance that all drivers could win?

1 Car / 20 places = 1/20 chance he could win

Since it is a team sport, what is the probability that either of the Red Bull cars could win?

2 Cars / 20 places = 1/10

What is the probability that the cars will not win where winning is equal among the others???
Using the M x N rule, we can say
(19/20) * (19/20) = 361/400 = 0.903

Since there are 20 cars on the track, there are so many permutations that could happen in terms of order with placing first second, or third.

$$P_{20}^{3} = \frac{20!}{(20-3)!} = 6840$$

There are 6840 possible outcomes for 20 cars to place on the podium (1st,2nd, or 3rd) in a given race.

Since being on the podium in itself is very good regardless of winning, the amount of combinations for just being on the podium is 1140

$$C_{20}^{3} = \frac{20!}{3!(20-3)!} = 1140$$

Conditional Probability

Max Verstappen's winning percentage from 2015 to 2022 is
35 wins out of 184 races in that time period
His win percentage is 19.02%

Out of the 184 races in that time period, he placed in the top 5 104 times
104 top 5's out of 184 races is 56.52%

What is the conditional probability that Max does not win while in the top 5???
Event A = Max does not win
Event B = Max places in the top 5

$$P(A) = 100 - 19.02 = 80.98$$

$$P(B) = 56.52$$

$$P(Loses|InTop5) = \frac{.3750}{.5652} = .6635$$

There is a 66.35% chance that Max will lose given he is in the top 5

P(A) and P(B) are dependent variables because
P(A|B) != P(A)

B(B|A) != P(B)

P(A AND B) != P(A)P(B)

If Those hold, then the variables would be independent

BAYES HERE?

The probability mass function of Max Verstappen Winning is

$$P(x) = \{ \substack{X=Win \\ X=Loss} \; \substack{19.02 \\ 80.98} \text{ otherwise } 0$$

Probability Distributions

Binomial Distribution

From 2015 - 2022, for Red Bull, there has been a 17.38% chance of a DNF(Did Not Finish). This means that either the car had trouble or the driver crashed, not allowing the car to finish the race. If we divide that percentage by 2, the percentage for Max Verstappen having a DNF is 8.69

Out of the 184 races, 8.69% of Max's races were DNFs. What is the probability Max will have 3 DNFs in those races in that time period of 8 years?
p = .0869
q = .9131
n = 184
y=3

$$P(y) = \binom{184}{3}.0869^3.9131^{184-3} = .07799$$

There is a 7.799% chance that Max will have 3 DNFs in that time period based on the binomial distribution

Geometric Distribution

With this information in mind, It is very likely that Max Verstappen will finish the race. In the current season, Max has 5 races left in the season and needs to finish 4 to have enough points to be the season champion. What is the probability that he will have an engine failure on or before the 5th?
p=.0869
n=5

$$P(x \leq n) = 1 - (1 - .0869)^5 = .3652$$

Using the formula from class, the chance that Verstappen will have an engine failure on or before the 5th/last race in the season is 36.52%

Hypergeometric Distribution

Out of the 20 Racers on the grid, we want to grab 5 random drivers to interview about their point of view in the race. We want a diverse perspective of drivers from different teams. What is the probability of the 20 randomly selected drivers we get Max Verstappen and Sergio Perez?

N = 20
r=2
n = 5
y=2

$$P(x=2) = \frac{\binom{2}{2}\binom{5-2}{20-2}}{\binom{5}{20}} = .05263$$

The probability that both drivers will be Red Bull drivers is 5.263%

Negative Binomial Distribution

Max Verstappen has a win rate of 19.02 percent. The Monaco Grand Prix is a historic event in formula one and is one of the most popular races among rich people with yachts. Winning in Monaco is a great accomplishment in Formula 1. Max has 97 wins under his belt. If there are 7 races till the Monaco Grand Prix, what is the probability that Max Verstappen will win 2 races before then, allowing him to potentially have his 100th win there?
p=.1902
q=.8098
r=2
y=7

$$P(x=7) = \binom{7-1}{2-1} * 0.1902^2 * .8098^{7-2} = .0755$$

There is a 7.55% chance that Max can have the chance of getting his 100th win at Monaco.

Poisson Distribution

Max Verstappen gets an average of 4 wins per season. What is the probability that he will win 10 races in the next season? I left last year out because he won 19 races in one season and I wanted to see the percentage of the outcome of the 2023 season. Really interesting!

$$P(X = 10) = \frac{4^{10} * e^{-4}}{10!} = .00529$$

There is a .529% chance that Max Verstappen will get 10 wins in the next season.

Tchebysheff's Theorem

Max Verstappen has an average finishing position of 4.69375 and a standard deviation of 2.544. What percentage of values fall between 5th and 10th place?

lower bound:
|4.69375-5|=0.30625

Upper bound:
|4.69375-10|=5.30625

k = 2.085

$$P(5 \leq X \leq 10) = 1 - .2299$$
$$P(5 \leq X \leq 10) = 0.7701$$

About 77.01% of the values fall between 5th and 10th place

Uniform Distribution

Max Verstappen's lap times have a uniform distribution of the interval of 100-180 seconds. What is the probability that Max completes a lap in less than 140 seconds in this lap?

$$f(x) = \frac{1}{180 - 100}$$
$$f(x) = .0125$$
$$P(X < 140) = \int_{100}^{140} \frac{1}{80} dx$$
$$P(X < 140) = 1.75 - 1.25 = .5$$

There is a 50% chance that max will complete a lap time in less than 140 seconds.

Continuous Random Variable

Given *f*(*y*)=*cy*2,0≤*y*≤2,and *f*(*y*)=0 elsewhere,find the value of *c* for which

f (*y*) is a valid density function.

In Max Verstappen's race Y is the number of of pit stops he takes. We want to find the value V in which f(y) is a probability density function. The number of pit stops per race at this particular circuit is limited from 0 to 2.

Y=0 is no pit stop

Y=1 is 1 pit stop

Y=2 is 2 it stops

$$given\ f(y) = vy^2 for 0 \leq y \leq 2\ and\ f(y) = 0\ otherwise$$

$$\int_0^2 cy^2 dy = c[\frac{y^3}{3}]_0^2 = \frac{3}{8}$$

Discrete Bivariate Distributions

Of nine drivers in a formula 1, four are rookies, three have never won, and two were world champions. Three of the racers are to be selected for special promotion. Let *Y*1 denote the number of rookie executives and *Y*2 denote the number of never won executives among the three selected for the award. Assuming that the three are randomly selected from the nine available, find the joint probability function of *Y*1 and *Y*2.

$$\mathrm{P}(\mathrm{Y}_1 = \mathrm{y}_1, \mathrm{Y}_2 = \mathrm{y}_2) = \frac{1}{84}\left[\binom{3}{y_1} \times \binom{4}{y_2} \times \binom{2}{3 - y_1 - y_2}\right]$$

	y1 = 0	y1 = 1	y1 =2	y1 = 3
y2 = 0	0	3/34	6/84	1/84
y2 = 1	4/84	24/84	12/84	0
y2 = 2	12/84	18/84	0	0
y2 = 3	4/84	0	0	0

After finding the joint probability function, an engineer decided to find the marginal distribution of y1

	y1 = 0	y1 = 1	y1 =2	y1 = 3	p(y1)
y2 = 0	0	3/34	6/84	1/84	10/84
y2 = 1	4/84	24/84	12/84	0	40/84
y2 = 2	12/84	18/84	0	0	30/84
y2 = 3	4/84	0	0	0	4/84
p(y2)	20/84	45/84	18/84	1/84	1

The marginal probability distribution of y1 is highlighted in red.

As the data shows, even one of the best drivers may not be as consistent as we once thought. The more data we have, the more we can see exactly how good of a driver he is. Even though Max Verstappen was not the best driver starting out, he got much better as time went on and became more consistent. In Formula 1, analysts are very busy with reading the information. During every race, back at their headquarters, they have a room similar to control station for NASA, but they have supercomputers running simulations of different outcomes for the race. There is more to the racing than driving the car, there are strategies involved and with the right strategy a slower driver could out perform the fastest driver. There are so many variables involved. Nonetheless, Max Verstappen is an extremely talented, consistent driver the more he competes.

Sources

http://www.espn.com/racing/driver/stats/_/id/4665/max-verstappen

https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020

https://www.motorsportstats.com/driver/max-verstappen/summary/series/fia-formula-one-world-championship

https://www.instagram.com/f1statsguru/?hl=en