

Python Project

Purpose:

As someone who practices both kick boxing and taekwondo as a hobby, I'm a big fan of mixed martial arts. The UFC (Ultimate Fighting Championship) is the world's biggest mixed martial arts promotion.

For those unfamiliar, mixed martial arts is a sport where two people fight in a cage. They can strike or grapple with their opponents. You can win by knockout, submission (Where one fighter gives up because they are in a choke or some other submission hold.) or decision (The judges score cards if the fight had no finish.)

I wanted to answer some questions I was curious about professional mixed martial arts fights using python and data scraped from the UFC stats site about their fighters. It contained data about the last 10 years' worth of fights.

Questions I would like to answer:

1. Do southpaw fighters have a significantly higher win rate against orthodox fighters?

Note: An orthodox stance is a stance with your left leg and hand in front typically used by right-handed people because in the orthodox stance your right hand is your power punch. Conversely South Paw is typically used by left-handed people and has a lead right leg and hand.

2. How often does a finish happen and how so? Does it change with weight classes?

3. Does having a longer reach than your opponent significantly boosts your chances of winning?

4. Does age affect win rate? What age range has the highest win rate?

Hypothesis:

1. Yes South Paw fighters will have a higher win rate. This is because most fighters are orthodox so a south paw fighter will get more practice against an orthodox fighter than vice versa.

2. Unsure but possibly knock outs happen more in higher weight classes due to punches being heavier.

3. Yes, more reach means you can hit your opponent easier.

4. Yes, old fighters are slower and weaker than when they were young. Conversely very young fighters often lack experience, so maybe somewhere in between. i.e. 30 years old has the highest win rate.

Data analysis:

To investigate if southpaw fighters had a significantly higher win rate against orthodox fighters, I looked at all southpaw versus orthodox match ups. There were 1444 such match ups. Out of those southpaw fighters won 745 fights. This was a win rate of 51.59%.

If we assume the null hypothesis is that there is no advantage in stance, then all else equal the probability of winning should be 50% for southpaw. To check if this 51.59% win rate is likely to be just due to variance or chance, I conducted a P test.

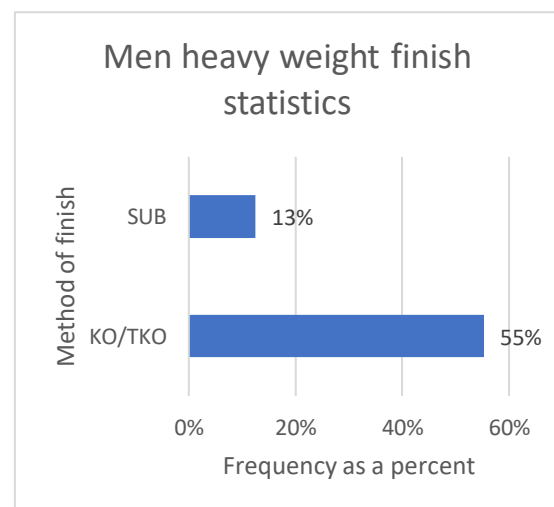
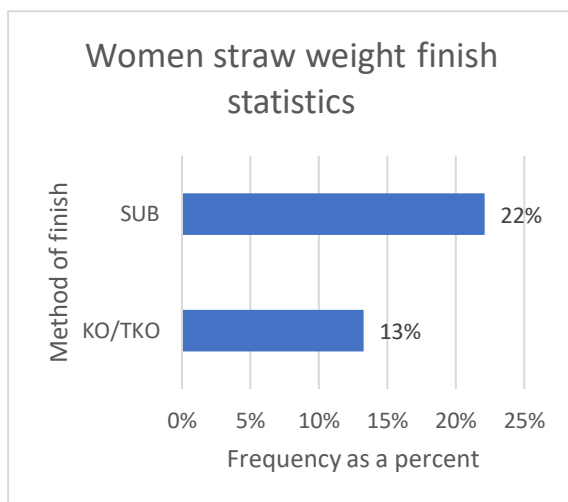
N was 1444 as we had 1444 fights, p was 0.5 as it should be 50% if there was no advantage in southpaw and the alternative was that p is greater than 50% due to southpaw having an advantage. I chose the P threshold as 0.05.

The P value yielded was 0.118 meaning there was an 11.8% chance of the observed result if I assume the null hypothesis is true, as this was higher than 0.05, I did not reject the null hypothesis.

In the second question, I wanted to see how often a finish occurred. I also wanted to see the types of finishes and if it varied with weight class.

My findings were that 30.5% of fights were won by knock outs, 17.4% by submission and the rest by decision with the occasional rare disqualification.

However, the weight class changed the likelihood of a finish and type of finish. For example, in the charts below when comparing men's heavy weight with women straw weight, there are big differences in how fights end.



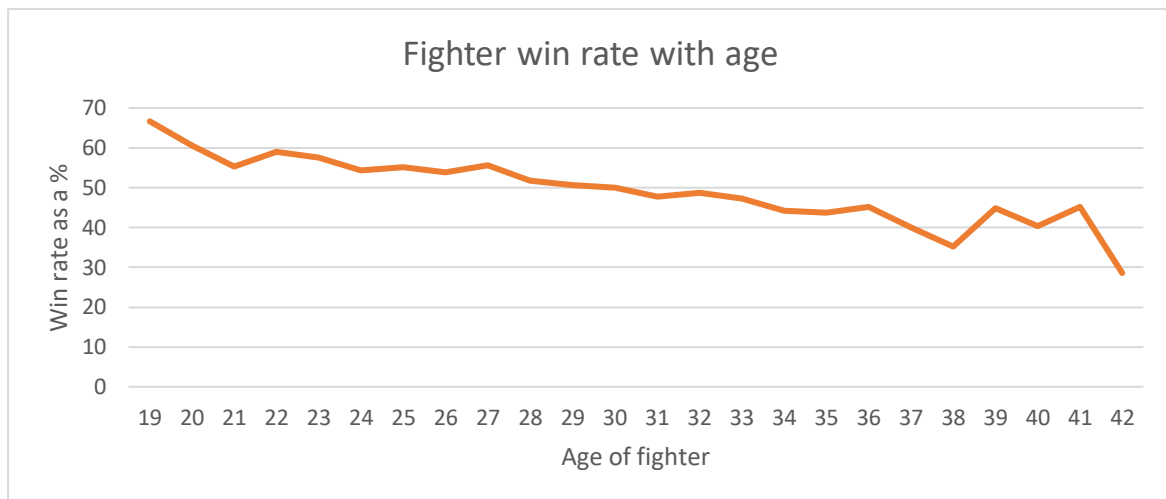
My third question was if having a reach advantage leads to a higher win rate. I looked at all fights where one side had a reach advantage. There were 4286 fights and the side with the reach advantage won 2261 of them. That is a win rate of approximately 52.8%.

Similarly, to my first question, I conducted another P test. I assumed the null hypothesis probability to be 50% and the alternative hypothesis to be that there is a higher probability of winning with a reach advantage. I used a threshold of 0.05.

The P value was 0.00016 which is lower than 0.05. In this case I rejected the null hypothesis, so I concluded that a reach advantage does correlate with higher win rates.

For my final question, I wanted to see how age affected win rate. I took all the wins grouped by age and divided it by all the fights group by age to find the win rate of each age.

My findings were that typically the win rate decreased with age. This is shown in the line chart below.



Result:

1. Southpaws had over 50% win rate against orthodox fighters. However, given the P value was greater than 0.05, it is not conclusive to say being a southpaw fighter will increase your chances against an orthodox fighter.
2. Yes, weight classes greatly affect the chances of a knockout. Heavier weight classes hit harder leading to an increased chance of a knockout.
3. Yes, a reach advantage is correlated with a higher win rate.
4. According to the data, win rate decreases with age. This differs from my hypothesis that the age with the highest win rate would be around 30 years old.