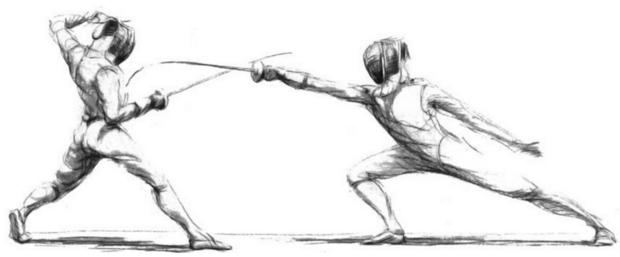
Exploration of Épée Fencing at the Tokyo Olympics

Oscar Fawcett, Abigail Hatcher, Simran Sidhu, and Skyler Won

October 2021



Introduction

Fencing as a sport traces all the way back to the 14th century, but its origins are rooted in the deadly duels from earlier ages. This may be why the fencing community today values good sportsmanship and proper protective gear more highly than other professional sports. Fencing is a sport where two players engage in sword combat on what is known as a piste with dimensions of approximately 14 meters by 2 meters. Every time a player touches their opponent with their sword in the appropriate area of the body, they score a point. The first player to 15 points or whoever has the most points after 3 three minute bouts wins the match. There are three main branches of fencing: sabre, foil, and épée. Each fencing type has its own variations and rules, but all have the same basic premise laid out above. This analysis will focus completely on épée fencing. We chose épée over sabre and foil because the whole body is considered a valid target as long as the point of the sword makes full contact. Because of this, épée fencing is slower and more precise than in sabre and foil, and thus easier to collect data on. Another aspect of épée that is not present in foil and sabre is the double. A double occurs if a player hits their opponent within 40 milliseconds after being touched themselves. In the event of a double, both players score a point. For this project, we set ourselves out to explore two questions. First, what situations are most likely to result in a double? This question may seem innocuous, but if there is a factor that contributes to the likelihood of a double, it could assist players in figuring out what they need to do to prevent doubles from occurring if they're behind, or increase their chances of a double occurring if they're ahead. Second, do there appear to be differences between the men's and women's épée fencing divisions? Due to the increased frequency of mixed gender games in the sport, answering this question could give useful information with regards to tournament organizers. To analyze these questions, we recorded 11 variables across 10 different épée fencing games.

For our recordings, we decided to watch games from the 2020 Tokyo Olympic's individual épée matches hosted by NBC. Each group member was responsible for collecting game data on the point-scored level. Our data was recorded in a shared Google Sheets document, which consists of 225 total observations.

There were fifteen unique players analyzed with eleven different male players and four different female players. The following variables were recorded for each point of the game: *match, player1, player2, league, body, blade_contacts, double, area, time, period*, and *scorer*. The *match* variable holds what match is currently being observed. The *league* variable holds which division of épée fencing the match is from; 'M' for men's and 'W' for women's. The variables *player1* and *player2* each hold the name of one of the players in the match. The four variables just described are categorical inputs in order to distinguish between the different matches analyzed, the other seven variables were more so used to describe what led to get the point or how the point was scored. The *blade_contacts* variable is of type integer and holds the number of times the blades of the players touched before the point was scored. The *body* variable indicates where on the body the tip of the sword touched. For the ease of

Match	Player1	Player2	League	Body	Blade Contacts	Double	Area	Time	Period	Scorer
Gold Medal	Romain Cannone	Gergely Siklosi	M	arm	10	FALSE	opponent on guard line	144	1	Romain Cannone
Gold Medal	Romain Cannone	Gergely Siklosi	M	chest	4	FALSE	center	130	1	Romain Cannone
Gold Medal	Romain Cannone	Gergely Siklosi	M	chest	2	FALSE	center	118	1	Gergely Siklosi
Gold Medal	Romain Cannone	Gergely Siklosi	M	chest	4	TRUE	center	91	1	Gergely Siklosi
Gold Medal	Romain Cannone	Gergely Siklosi	M	chest	1	FALSE	center	77	1	Romain Cannone
Gold Medal	Romain Cannone	Gergely Siklosi	M	chest	10	FALSE	opponent on guard line	40	1	Gergely Siklosi

analysis, we restricted the *body* variable to be equal to 'head', 'chest', 'back', 'arm', or 'leg'. The *double* is a Boolean that holds true if a double occurred, and false otherwise. The *area* variable holds where on the piste the point was scored. A fencing piste can be broken up into 5 parts; the center, two on guard lines, and two warning lines which mark the last 2 meters of the piste on both ends. Given this, *area* can hold the values of 'center', 'on guard line', 'opponent on guard line', 'warning line', or 'opponent warning line'. In order to determine the value, we created an imaginary line in between the two players, then whichever area of the piste that contains the most length of that line was inputted in for *area*. The *time* variable holds how much time is left on the clock after the point was scored and the *period* variable holds which period the game is in, with values 1 through 3 representing periods 1 through 3 respectively and 4 representing overtime. Lastly, the variable *scorer* indicates which player made the point, or in the case of a double, who hit first. We also created a new variable *round_time* for our analysis from *time*, which holds the amount of time it took until the point of scored.

Summary

Table 1

Variables	Min	Q1	Median	Q3	Max	Mean	SD
blade_contacts	0	2	5	9	41	7.02	6.99
time	0.00	29.00	71.00	118.00	174.00	73.98	50.30
round_time	1.04	8.00	16.00	29.00	130.00	20.82	17.23
double	0	0	0	0	1	0.13	0.34

Table 1 contains the five number summary along with the mean and standard deviation for the three quantitative variables *blade_contacts*, *time*, and *round_time* and the Boolean variable *double*. Though it may not make much sense to calculate these statistics for the variable *time*, we did so for the sake of posterity.

Table 2

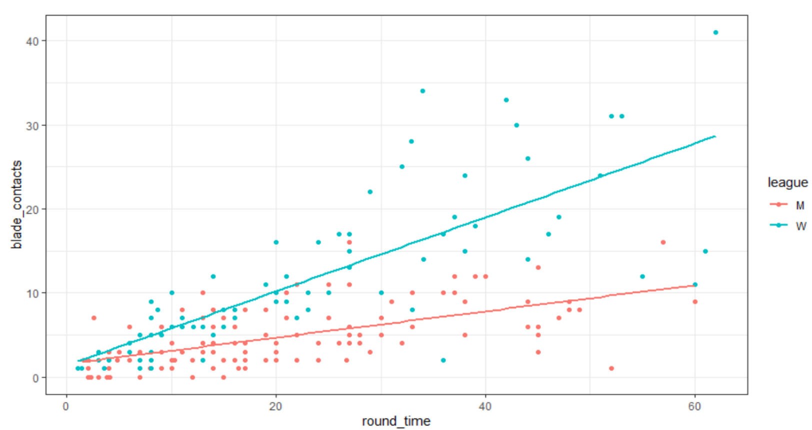
league	Count	%
M	143	0.64
W	82	0.36
Total	225	1.00

body	Count	%
arm	42	0.19
chest	130	0.58
head	1	0.00
leg	13	0.06
back	18	0.08
Total	204	1.00

area	Count	%
warning line	14	0.06
on guard line	52	0.23
center	100	0.44
opponent on guard line	53	0.24
opponent warning line	6	0.03
Total	225	1.00

Table 2 contains the frequency and relative frequency for the categorical variables *league*, *body*, and *area*. Note: the total number of observations for body is only 204. This is because if the round is stopped before a point was scored (such as when the time in the period runs out), the body variable holds the value of NA while the other variables can still be recorded as normal.

Figure 1

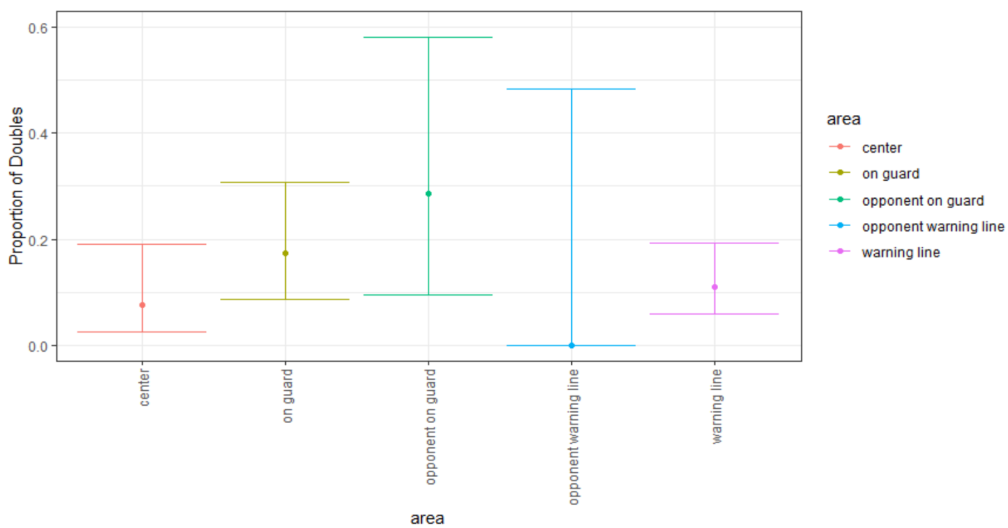


In figure 1, we can see the relationship between the variables *blade_contact* and *round_time* with *league* as a grouping variable. For this plot we removed two influential points with *round_time* values > 75. Notice how the the rate of change for the women's division is greater than that for the men's division. To put analytics behind what we are observing, we created the following linear regression model:

$$blade\_contacts = round\_time + round\_time : league$$

With league being 1 if in the women's division and 0 for the men's. For this model, we calculated a respectable R^2 value of 0.67 and a significant p-value for the coefficient of *round_time:league* of $2 * 10^{-16}$, meaning that we can say the slopes for the men's and women's divisions are different.

Figure 2



In figure 2, we can see the proportion of doubles that occurred in each area of the piste along with their 95% confidence interval. Notice how each confidence interval overlaps with one another. This seems to indicate that whether or not a double occurs is not correlated with the variable *area*.

Insights

Double Predictability

Similar to Figure 2, we wanted to see if the data we collected could be utilized to predict whether a double will occur, or strengthen our observation in Figure

2. Although the amount of doubles that occurred in any given match is low, as evident by Table 1, we decided to move forward. To answer this question, we first created 5 logistic regression models to predict whether or not a double occurred.

$$Model1 : double = body$$
$$Model2 : double = area$$
$$Model3 : double = blade\_contacts$$
$$Model4 : double = round\_time + period$$
$$Model5 : double = body + area + blade\_contacts + round\_time + period$$

The 1st model's goal was to observe whether an occurrence of a double could be predicted based on the area of the body the point was scored. However, when running the model, the p-values were extremely high, ranging from 0.985 to 1, indicating that using the body variables to predict whether or not a double occurs is essentially random.

body	p-value
intercept	0.985
back	0.986
chest	0.987
head	1.000
leg	0.987

Our 2nd logistic regression model using area provided similar results with high p-values. However, we did see a p-value of 0.107 for the warning line area but we can attribute this to the low amount of data points on the warning line.

The 3rd and 4th models results were comparable with our previous two models. With P-values of 0.77 for *blade_contacts*, 0.759 for *round_time*, and 0.463 for *period*. These results are evidence that *blade_contacts*, and *round_time + period* have no correlation with doubles.

By generating a full logistic regression model with all our recorded variables, we continue to see no statistically significant results. We did observe one statistically significant result in 'warning line' for the variable *area* with a p-value of 0.0104, but we don't think this should be of any particular note. While the p-value may be low, we recorded very few cases of points being scored on the warning lines and even fewer cases of doubles, thus we have to look at this model with skepticism.

These findings led us to the general conclusion that there is no relationship between an occurrence of a double and the variables we recorded.

Differences Between the Men's and Women's Leagues

Based on figure 1, we observed a difference in the distribution of the men's league vs the women's league. From this discrepancy, we took a deeper look into the differences between the two leagues.

We begun this by first investigating some of the information present in figure 1. We did this by creating 95% confidence intervals for the mean of the variable *blade_contacts* across the men's and women's divisions. This led us to the following results:

league	Lower Bound	Mean	Upper Bound
M	3.867	4.441	5.014
W	9.443	11.439	13.435

Since the intervals of the men's and women's divisions do not overlap, we can say that the average number of blade contacts per round is different across the leagues.

The variable *round_time* was also investigated in the same manner, but in this case, though there did appear to be a potential difference with the women's division having higher on average *round_time*, the results were insignificant.

Finally, we created a new metric called engagement to measure how "engaged" or active the fencers were in a match. By using this new metric we could estimate how fast paced the set was. The engagement was calculated as:

$$engagement = \frac{blade\_contacts}{round\_time}$$

Blade contact and round time on their own do not give a good indication of how a particular session went. However, by utilizing both variables, we can attempt to piece together the general flow of each league. For example, high blade count and a short round time could mean a very high-action round.

Once again, we constructed 95% confidence intervals for each respective league.

engagement	Lower Bound	Mean	Upper Bound
M	0.2616	0.3138	0.3660
W	0.4768	0.5271	0.5774

From this confidence interval, we can establish that the women's épée fencing league has a higher average amount of engagement per bout. Using the insights and observations we could glean from our figure and confidence intervals, we can come to the conclusion that Women's épée fencing matches have a faster-paced game with more engagement while having similar round times with the Men's league.

Moving Forward

Looking back on our data, though we stand by the analyses performed in this project, there are a few places for improvement. The implications of this project are limited in scope due to the fact that we collected data solely on the Tokyo Olympic Games. Therefore,the results of this project can only be applied thus far to the highest echelons of épée fencing play. In addition to this, there are some sample size issues with our data. For starters, we don't have many observations where doubles occurred, nor when points were scored at the warning lines. Additionally, the players across recorded games were not always unique. As a result, our collected data may not be representative of all épée Olympic athletes.

There were also some issues with the collection of data. Due to the speed of the fencing athletes, it was difficult at times to figure out where a player's sword touched their opponent when they scored a point. The institution of a device that could collect that data via each player's fencing gear would provide a more accurate way to measure this variable. This device would also make the divisions between the parts of the fencer's body more precise. Similar to the vagueness of the *body* variable, we only collected data on where the players were on the piste at the moment a point was scored, but this variable doesn't take into account the movement that the players made during the bout. Using player tracking equipment to collect the positions of both players roughly every second would be a welcome addition into the fencing data science world. Overall, we are proud of the work done in this project. At the current moment, data analytics in the sport of fencing is hard to come by, but hopefully in the near future advancements in this field grow so that further investigation can occur into the questions we have laid forth.