

Row Row Row Your Boat

Math 114 Mathematical Modeling

St. Mary's College

March 8, 2019

Problem Statement

If you have been to a rowing regatta, you may have observed that the more oarsmen there are in a boat, the faster the boat travels. Investigate whether there is a mathematical relationship between the speed of a boat and the number of crew members. Consider the following assumptions (partial list) in formulating a model:

1. The total force exerted by the crew is constant for a particular crew throughout the race.
2. The drag force experienced by the boat as it moves through the water is proportional to the square of the velocity times the wet surface area of the hull.
3. Work is defined as force times distance. Power is defined as work per unit time.

Hint: When additional oarsmen are added to a shell, it is not obvious whether the amount of force is proportional to the number in the crew or the amount of power is proportional to the number in the crew. Which assumption appears the most reasonable? Which yields a more accurate model?

Data

Crew	Distance	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6
1	2500	20:53	22:21	22:49	26:52	0	0
2	2500	19:11	19:17	20:02	0	0	0
4	2500	16:05	16:42	16:43	16:47	16:51	17:25
8	2500	9:19	9:29	9:49	9:51	10:21	10:33

Distance is in meters and time is in minutes:seconds

Commentary

For this problem we are deciding to make a model that will return the race time of a team given how many people are in the boat and what race the team is currently running. Looking at the data, we can see that time increases as the number of races performed increases. This suggests that those that are participating in the race are getting more tired as they participate in more races. Depending on the size of the team, different energy systems of the teammembers are being exhausted to a different degree. Due to the lack of data, we will assume that teams of size 1-2 tire in a way that is different than teams of size 4-8. With respect to the measure of force vs power, the measure of force will not be useful when trying to model the fatigue of teams, as the amount of force required to complete the race has not changed, even though the time will. If we use power, we will be better able to see how the team fatigued as they participated in consecutive races.

Given this discussion, we will be transforming the data above from time to power. We will create two models, one that predicts the first race time, and another that predicts the rate at which the teams will fatigue. Our model will support team sizes of 1,2,4,8 and races numbers of 1-6.

Power Data

We are told from our list of assumptions that power is work per unit time, work is force times distance, and the force acted on the boat in order to move it is $force = k * v^2 * SA_{hull}$, $v = distance/time$. We can reduce the equation in the following way:

$$power = work/time$$

$$power = (force * distance)/time$$

$$power = ((k * v^2 * SA_{hull}) * distance)/time$$

$$power = ((k * (distance/time)^2 * SA_{hull}) * distance)/time$$

$$power = ((k * (distance^2/time^2) * SA_{hull}) * distance)/time$$

$$power = ((k * (distance^3) * SA_{hull}))/time^3$$

$$power = \frac{k * distance^3 * SA_{hull}}{time^3}$$

For the sake of simplicity, we will allow k to be one. Also for the sake of simplicity, we will allow SA_{hull} to be equal to the number of people in the boat. To actually find k , we would most likely need more data. To get a more accurate value of SA_{hull} , we'd have to get the size of the boat used for a crew of size 1,2,4,8, as the boat sizes are not constant, as well as the data needed to calculate how far below the water the boat would be, which would require some more physics as well as data to properly predict. Setting $k = 1$ and $SA_{hull} =$ the number of people in the boat, we get the following data:

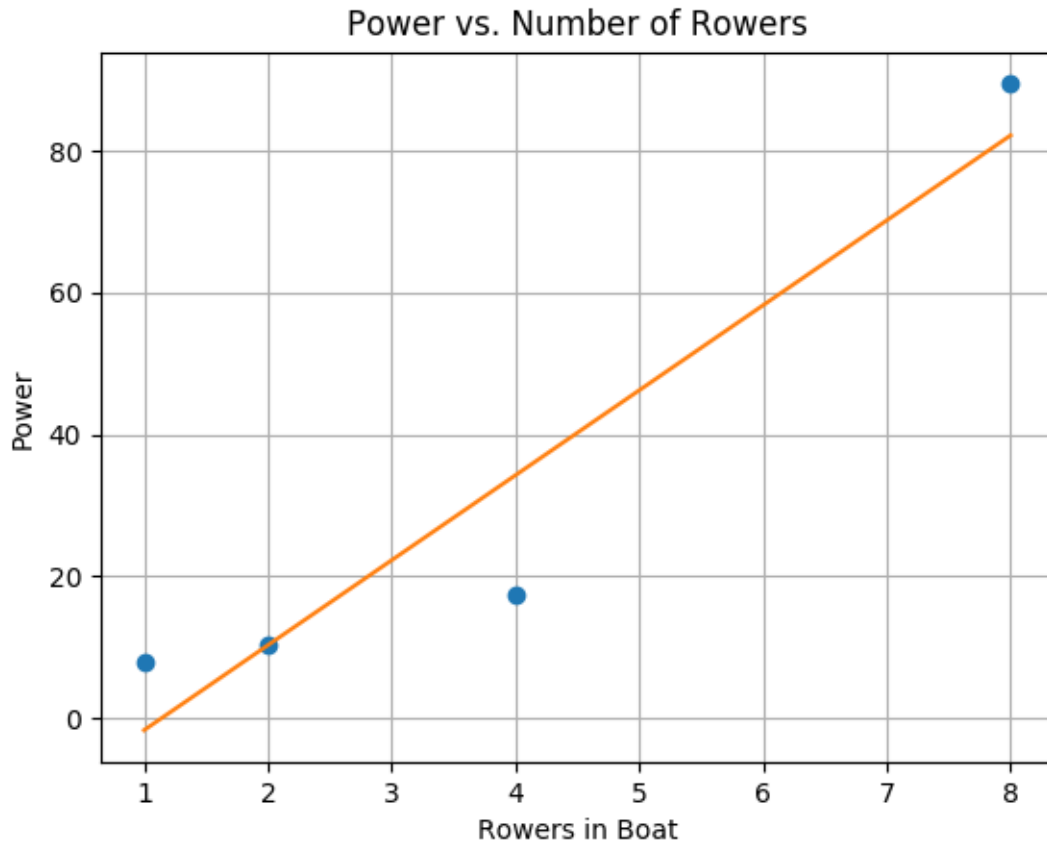
Power Data

Crew	Distance	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6
1	2500	7.94268	6.47938	6.08989	3.73014	0	0
2	2500	20.4939	20.1767	17.9944	0	0	0
4	2500	69.5502	62.1265	61.9409	61.2057	60.4821	54.7685
8	2500	715.607	678.536	611.736	605.547	521.958	492.832

Model 1: The first race

The first model that we are creating will output how much power a crew will exert during a race given the number of people in the boat. Since most of the data we have is based upon subsequent races, we will only be using data from race 1. When thinking about what kind of model we should fit to our data, we should consider the power that each rower brings to the team. For the sake of simplicity, we may assume that all rowers output the same power, and that the only people in the boat are actively rowing. This means that none of the people in the boat are acting as the coxswain, a position in boats of larger teams whose responsibility it is to steer the boat and set the tempo instead of doing any actual rowing. Following these assumptions, we assume that the power output of every athlete is the same and that every athlete tires at a uniform rate. Given these assumptions, we can conclude that every additional athlete increases the boat's ability to output power by a constant number (e.g power output of boat with n rowers $= P_n$, $P_{n+1} = P_n + C$). While other regression techniques exist that yield a smaller least squares value, a linear regression most aptly matches our assumptions.

Using polyfit from the library numpy, we get $104.383x - 188.038$, which has a least squares of 604406.716.



Model 2: Fatigue

Our goal for this model is to calculate how an individual tires as they consecutively race, and then use the race number, as well as the number of people in the boat to determine how much slower the team will perform as they continue to race. In order to calculate how a person slows as they race we will look at the difference in power outputs at a certain race. We will then divide this difference by the number of people in the boat, and then average these values to determine how much of a decrease we might expect for a given race number.

$$Diff(x) = (Race_{x-1} - Race_x) / (SizeofCrew)$$

Where the first difference is equal to the first race time minus the second race time

divided by the number of people in the boat. Diff 1, People in Boat 8, for example, is equal to $(678.536 - 715.607)/8 = -4.634$.

Fatigue Data

People in Boat	Diff 1	Diff 2	Diff 3	Diff 4	Diff 5
1	-1.463292	-0.389490	-2.359755	NaN	NaN
2	-0.158591	-1.091161	NaN	NaN	NaN
4	-1.855914	-0.046409	-0.183799	-0.180902	-1.42838
8	-4.633814	-8.350047	-0.773691	-10.448627	-3.64070

Looking at the data, a boat with 8 people has a very different diff than any other boat. This leads me to hypothesize about the fluidity of the rower's strokes in the boat of 8 people. If the uniformity of the rower's stroke affects the race time, and a rower might struggle to operate in uniformity if they are tired, and it is harder for multiple rowers to row in uniformity, especially if they are tired, then the 8-person boat might have such a high diff when compared to boats that have less people in them. Due to lack of data, this can neither be proven nor disproven, so we will use the data despite the fact that it is visibly different than the other boats.

Fatigue to Expect

After 1	After 2	After 3	After 4	After 5
-2.028	-2.469	-0.829	-2.657	-1.267

Since the force that is applied is constant, we can check how good our approximation is using least squares on the force data. We now take the number of people in the boat, and what race they are running, and use this to predict the power output for their race. This is

done with the linear regression that was performed earlier in the project. Depending on race that the rowers are on, we then subtract the necessary fatigue constant from that result. If they are on race 2, we subtract 2.028, if they are on race 3, we subtract 2.469, etc. To summarize, we have the following equation:

$$a = 104.38298143764881$$

$$b = -188.03775163057534$$

$$p(x, r) = ax + b - F_r$$

Where $p(x, r)$ is the expected power of a boat of x rowers after r races.

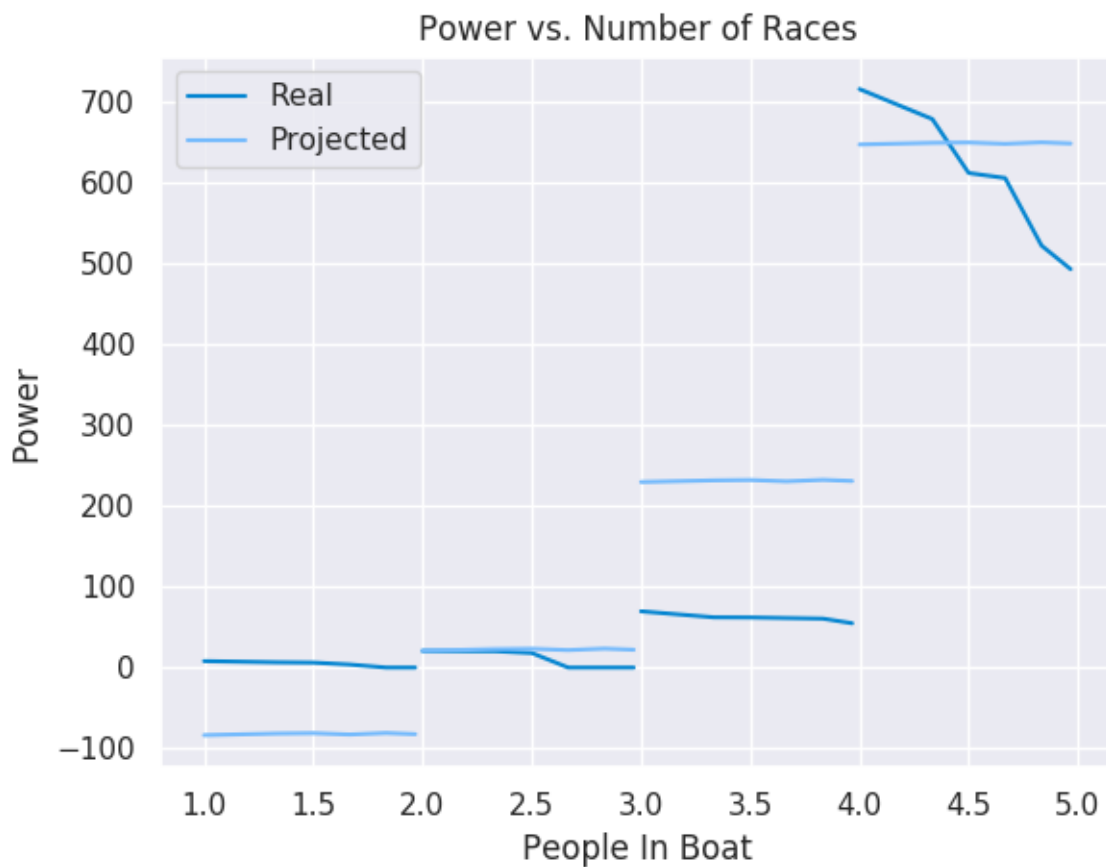
This modeling strategy returns a least squares of 267639.0378 and the following tables and visualization.

Actual Performance

Crew	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6
1	7.94268	6.47938	6.08989	3.73014	0	0
2	20.4939	20.1767	17.9944	0	0	0
4	69.5502	62.1265	61.9409	61.2057	60.4821	54.7685
8	715.607	678.536	611.736	605.547	521.958	492.832

Predicted Performance

Crew	Race 1	Race 2	Race 3	Race 4	Race 5	Race 6
1	-83.654770	-81.626867	-81.185493	-82.825459	-80.997388	-82.387500
2	20.728211	22.756114	23.197488	21.557522	23.385593	21.995481
4	229.494174	231.522077	231.963451	230.323485	232.151556	230.761444
8	647.026100	649.054003	649.495377	647.855411	649.683482	648.293370



This visualization is purposefully confusing. The number of people in the boat doesn't increase, the teams just participate in new races. This was done to properly visualize how the power output of the teams is affected by the participation in consecutive races.

Conclusion

Though the effectiveness of the modeling was hamstrung by the lack of data, this project does a very good job of showing the different assumptions and thought processes that are used when walking through a data science problem. If I did this project again, I would use an exponential curve to model the expected power output and I would make a different set of fatigue constants for every race. This would result in a much better model, however there is a good chance that this model would suffer from overfitting. If there was more data, I would love to test my assumptions about fatigue and power output given the number of people in the boat. Looking forward to more complicated, data-heavy projects.