Moneyballing Rowing

Yih-Jen Ku APMA 4903 October 14, 2015

Outline

- The Problem - What is the fastest boat?
- References & Resources
- Background of the Data
- Rowing Data Analysis using Python
- External Factors and Future Outlook
- About Me

Outline

- **The Problem - What is the fastest boat?**
- References & Resources
- Background of the Data
- Rowing Data Analysis using Python
- External Factors and Future Outlook
- About Me

The Problem - What is the fastest boat?

- In collegiate rowing, boats are typically made up of **8 rowers**.
- The top boat is called the **varsity eight**. Each year, the team with the fastest varsity eight over **2,000 meters** wins the **national championship**.
- Given that a competitive Division I team usually has anywhere from 30 to 60 rowers, how can coaches determine the **fastest lineup** given the viable candidates?
- Using **linear regression** and **optimization**, we can try to predict the fastest lineup.

Quick Visual of Rowing

https://youtu.be/ovEsA5p83T0?t=2m

Why I Care

- Varsity coxswain of the lightweight rowing team at Columbia
- Coxswain responsibility: data collection and analysis
  - How can we effectively use this data?
- Help the team toward a national championship

Why You Should Care

- Strong athletics at Columbia = $$$ for Columbia and better events for students (like Homecoming- football finally won!!)
- Regression has many applications
  - Widely used in all fields such as medicine, business, finance, sports, etc. to find possible relationships between variables

Outline

- The Problem - What is the fastest boat?
- **References & Resources**
- Background of the Data
- Rowing Data Analysis using Python
- External Factors and Future Outlook
- About Me

References & Resources

General Overview of Concepts:

- "Moneyball." *Wikipedia*. Wikimedia Foundation, n.d. Web. 13 Oct. 2015. <https://en.wikipedia.org/wiki/Moneyball>.
- "Linear Regression." *Wikipedia*. Wikimedia Foundation, n.d. Web. 13 Oct. 2015.
- "Regularization (mathematics)." *Wikipedia*. Wikimedia Foundation, n.d. Web. 13 Oct. 2015.
- "Constrained Optimization." *Wikipedia*. Wikimedia Foundation, n.d. Web. 13 Oct. 2015.

Physiology and Rowing Research:

- Rowing Canada Aviron. "Rowing Athlete Development and Ranking." (2014): 1-11. Rowing Canada Aviron, Web. <http://www.rowingcanada.org/sites/default/files/athletemonitoringrcanov152014.pdf>.
- Mann, T. N., Lamberts, R. P., & Lambert, M. I. (2014). High Responders and Low Responders: Factors Associated with Individual Variation in Response to Standardized Training. *Sports Medicine, 44*(8), 1113-1124.
- Hagerman, F. C. (2000). Physiology of competitive rowing. *Exercise and sport science*, 843-873.

Python Resources:

- “Introduction to Linear Regression” adapted from James Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. iPython Notebook.
- Cournapeau, David. "1.1. Generalized Linear Models¶." *1.1. Generalized Linear Models — Scikit-learn 0.16.1 Documentation*. Web. 13 Oct. 2015. <http://scikit-learn.org/stable/modules/linear_model.html>.
- Johnson, Connor. "Linear Regression with Python." *Connor-johnson.com*. N.p., 18 Feb. 2014. Web. <http://connor-johnson.com/2014/02/18/linear-regression-with-python/>.

Outline

- The Problem - What is the fastest boat?
- References & Resources
- **Background of the Data**
- Rowing Data Analysis using Python
- External Factors and Future Outlook
- About Me

Moneyball: The Art of Winning an Unfair Game
by Michael Lewis

- An evidence and sabermetric based approach to assembling a baseball team with a limited amount of funds
- Looks at on-base percentage and earned run average over stolen bases, runs batted in, and batting average as **attributes** of offensive success
- Using these attributes, run **multiple regression** to evaluate pitchers and hitters in order to create a model that will predict runs scored, runs allowed, and overall wins.

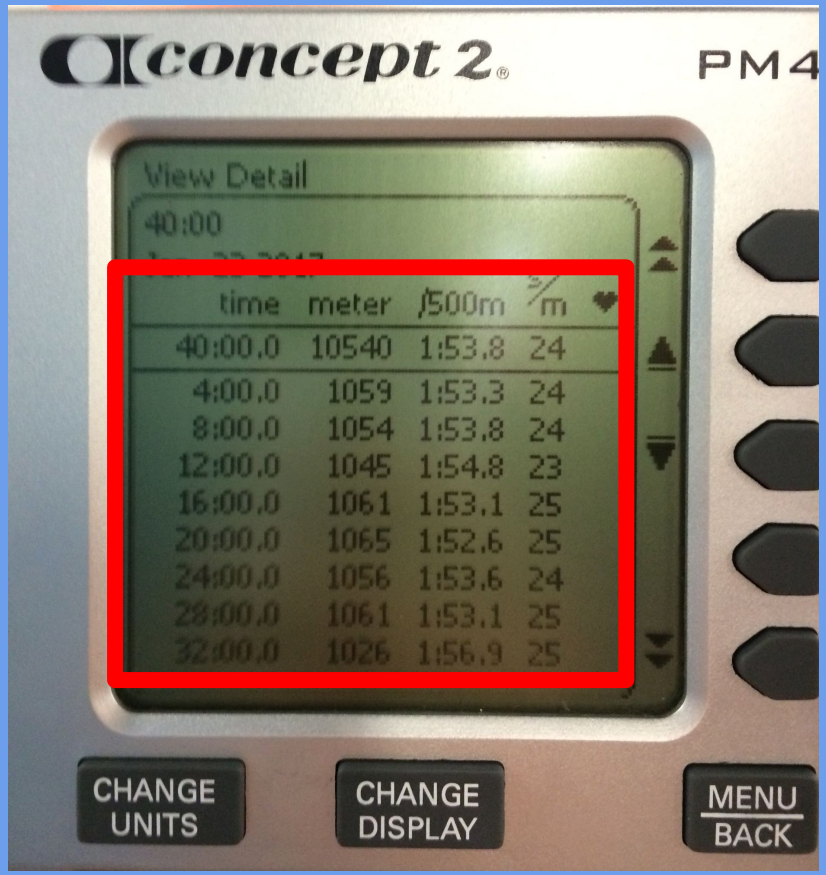
Statistical Attributes of Rowing

What are the measures of a rower's ability?

- Endurance - Forty Minute Test
  - meters traveled in 40 minutes
- Sprinting - One Minute Test
  - meters traveled in 1 minute
- Strength - Rep Max Test
  - average of max squat measure and max deadlift measure
- Power - Max Wattage Test
  - maximum wattage output
- Weight
  - measure weight in lbs
- ***Use these 5 attributes to predict the velocity of the boat***

Where do we get this data from?

Ergometers
(indoor rowing machines)



	Name	40 Minute Distance (meters)	1 Minute Distance (meters)	Rep Max (kg)	Max Watt (watts)	Weight (lbs)
1	Monahan	11173	344	259	676	154
2	Rivkin	10454	335	250	702	145
3	Ptucha	11106	356	338	812	154
4	Docter	10779	342	293	787	156
5	Blair	10708	346	323	796	154
6	Ross	11096	333	278	731	160
7	Murphy	10386	335	259	717	155
8	Hou	10780	350	320	791	154
9	Puttmann	10998	348	285	800	155
10	Smith	10591	328	262	615	155
11	Bae	10088	337	291	727	152
12	Solberg	11510	348	298	744	155
13	Simon	10986	353	245	775	160
14	Grueterich	10715	341	293	784	157
15	Mcgrattan	10808	353	308	800	152
16	Bartholomew	10112	349	275	743	157
17	Madani	10850	338	268	646	153

Individual rower attributes - use these to find the average for each lineup

Time Trials

Real results from a practice last March

- **20 time trials** each with a lineup of rowers
- We can calculate **20 velocities** (distance over time)
- Find the **average** for each attribute from the individual attribute values of the rowers
- Use the **velocity** and the **average attribute values** in the regression model

AM

Piece 1

1	Monahan	1	Puttmann
2	Rivkin	2	Smith
3	Ptucha	3	Bae
4	Docter	4	Solberg
5	Blair	5	Simon
6	Ross	6	Grueterich
7	Murphy	7	McGrattan
8	Hou	8	Bartholomew
	04:31.9		04:35.8

Piece 2

1	Monahan	1	Puttmann
2	Bartholomew	2	Smith
3	Ptucha	3	Bae
4	Docter	4	Solberg
5	Blair	5	Simon
6	Ross	6	Grueterich
7	Murphy	7	McGrattan
8	Hou	8	Rivkin
	04:31.5		04:37.4

PM

Set 1

1	Monahan	1	Puttmann	
2	Bartholomew	2	Smith	
3	Ptucha	3	Bae	
4	Docter	4	Madani	
5	Blair	5	Simon	
6	Ross	6	Rivkin	
7	Murphy	7	McGrattan	
8	Hou	8	Grueterich	Diff
	01:29.8		01:31.7	00:01.9
	01:35.6		01:37.7	00:02.1

Set 2

1	Monahan	1	Puttmann	
2	Bartholomew	2	Smith	
3	Ptucha	3	Bae	
4	Rivkin	4	Madani	
5	Blair	5	Simon	
6	Ross	6	Docter	
7	Murphy	7	McGrattan	
8	Hou	8	Grueterich	Diff
	01:32.1		01:33.6	00:01.5
	01:35.6		01:35.3	00:00.3

Set 3

1	Monahan	1	Puttmann	
2	Bartholomew	2	Smith	
3	Ptucha	3	Bae	
4	Grueterich	4	Madani	
5	Blair	5	Simon	
6	Ross	6	Docter	
7	Murphy	7	McGrattan	
8	Hou	8	Rivkin	Diff
	01:29.8		01:35.2	00:05.4
	01:38.3		01:39.1	00:00.8

Set 4

1	Monahan	1	Puttmann	
2	Docter	2	Smith	
3	Ptucha	3	Bae	
4	Grueterich	4	Madani	
5	Blair	5	Simon	
6	Ross	6	Bartholomew	
7	Murphy	7	McGrattan	
8	Hou	8	Rivkin	Diff
	01:30.6		01:35.3	00:04.7
	01:38.4		01:35.9	00:02.5

What is the model predicting?

Once we run the regression, it will return **5 coefficients** that we can use to calculate a velocity given 5 input variables (our 5 attributes).

Now let’s go to the code to see this in action!

Outline

- The Problem - What is the fastest boat?
- References & Resources
- Background of the Data
- **Rowing Data Analysis using Python**
- External Factors and Future Outlook
- About Me

Outline

- The Problem - What is the fastest boat?
- References & Resources
- Background of the Data
- Rowing Data Analysis using Python
- **External Factors and Future Outlook**
- About Me

External Factors

Are there variables out of our control affecting our model?

- “Ergs don’t float” - erg results don’t always translate well on the water
- Water and wind conditions affect the velocity of the boat
  - Not always doing time trials in same direction
  - Water and wind can change quickly
- Rower fatigue
  - Rowers will not be able to output the same amount of power for each piece since we do them back to back

Future Outlook

- How can we improve this model?
  - Refine attributes measuring the on-water ability of a rower
  - Use data from machines that better simulate on-water rowing
  - Take more data - 20 values probably isn't enough data for a sufficient regression model
- Research
  - My coach's research was primarily physiology-based which helps determine the attributes by which he measures our rowers
  - Would be interesting to speak with national team coaches and coxswains about how they interpret data
- Your thoughts?

Outline

- The Problem - What is the fastest boat?
- References & Resources
- Background of the Data
- Rowing Data Analysis using Python
- External Factors and Future Outlook
- **About Me**

About Me

- Majoring in Applied Math but also minoring in CS
- ADI Committee
- Goldman Sachs Technology
- Software Engineering
- Grad School in Data Science / Computer Science
- Continue coxing?