Kayak Final Performance Analysis  
Austin Schwinn  
Econometrics

The sport of flatwater kayaking is growing emensily in the US and around the world. From beginners going out on the lake for the first time to the most elite, top-level athletes; more people are joining on this exciting past-time. With this influx of participants, the competition level at the top tier of competitions is becoming fiercer. Within this high caliber population, it has become harder to stay at the top as more racers come up through the ranks. As an avid kayak racer, I wanted to use statistical analysis to see the likelihood to win one race right after another. On the surface level, flatwater kayaking seems to be a simple sport; whoever paddles from start to finish in the shortest amount of time wins. With this research I hope to see how true this basic sentiment is. My goal is to analyze the results from the 2012 Olympic Games to see how likely an athlete is to finish in the top 3 places, based how they finished in the heat and semi-final, as shaped by a number of different variables.

Before I started my regression and data analysis, I began with background research on the sport of kayaking and what could affect the outcome of their final race based on their previous performances. The first article I started with actually helped shape my research question. The article “Variability and Predictability of Elite Canoe-Kayak Performance” set out to see how the final placing in races are shaped by their semi-final run times. This article used linear modeling to analyze the times from the top finalist and how closely that relates to their position in the semi-finals. They eventually found that the variability between these two performances is high compared to similar sports. The difference between this research project and my own is the data we used to shape our analysis. Nibali, Hopkins, and Drinkwater used run times to see the fluctuation between semi-final and finals. I, on the hand, used places in the races. This is because race times in kayak sprints vary widely based on weather conditions, such as wind.

Another factor that I believe goes greatly into the final result of the races is a particular racers strategy. “Decision making in Elite White-Water Athletes Paddling on a Kayak Ergometer” found that at moderate activity, reaction time is sped up but it impairs direct response. If a kayaker is more likely to make a poor decision and stray from their race plan, there will be a higher fluctuation in race results. This can also be reflected in the article “The Influence of Pacing Strategy on VO2 and Supramaximal Kayak Performance” as it analyzes different passing strategies within the sport. This article found that an all-out plan gave a higher overall power output compared to a tamed, paced strategy. So an explosive lead off the start is better than trying to catch up at the end. As well as the race strategy, how much energy a racer uses throughout a regatta of heats, semi’s, and finals will affect their performance in the final race. “Gross Efficiency and Energy Expenditure in Kayak Ergometer Exercise” found that it is highly dependent on the efficiency of the particular paddler, their technique, and body type. The article “Morphological characteristics of Olympic slalom canoe and kayak paddlers” found that slalom kayakers had similar body types within the sport. This can be translated to Flatwater as well because it is mainly based on the power-weight ratio of the athlete. Since most of the top athletes of superb technique and fairly similar body-types, that shouldn’t have much of an effect on the final results.

With all this background information in play, I decided to merge my research question from these articles. I am trying to find what the likelihood of a racer finishing in the top 3, based on their previous finishes in the heats and semi-finals. These are also based on several variables such as gender, which heat and semi-final number they were in, how many people were in the boat, and what event distance the race was. Each variable is a factor that could affect the likelihood of finishing in the top of the final race, as you finish in precursor races.  
  
The data I used is from a results archive on sportsscene.tv, a website dedicated to the sport of kayaking. The results I used were from the spring kayak races that took place in the 2012 Olympic Games in London.

Data: See Appenix A-C

For my analysis of data, I broke it down in to 4 different regressions:

1) How likely are you to finish in the top 3 in the final based on your finishes in the heat and semi-finals, as well as what heat number and semi-final number you were in?

2) How likely are you to finish in the top 3 in the final based on your finishes in the heat and semi-finals, as well as the gender of the racer?

3) How likely are you to finish in the top 3 in the final based on your finishes in the heat and semi-finals, as well as how many people were in the boat (1, 2, or 4)?

4) How likely are you to finish in the top 3 in the final based on your finishes in the heat and semi-finals, as well as what event the race was (200m, 500m, 1000)?

Results: See Appendix D

I tested and found that my regressions are statistically significant, which means that all these factors go into how likely the racer is to finish where they do. The data shows that the later your heat and semi-final number are, the less variation there is in your performance. I believe this is because this gives the racer time to see the other racers go, letting their nerves ease. The regressions also show that there is a higher variation when the athlete is male, compared to the females. I would say this is due to the fact that the male pool of athletes is deeper and there tends to be more competition because it has had longer to mature. There is also more variation when there are more racers in the boat. I would say that this is because the more people there are in the boat, the more that can go wrong because there are more people that can vary how they feel. The last regression looked at the variance based on the event distance, in meters. It found that the longer the event, the greater the variance in the final place compared to the semi-final and heat. I believe that over a longer race, there is more that go wrong and there are many more variables that go into the race and race strategy.

I believe that this research will help expand on the research that was previously mentioned. One aspect that would help push this research further would to be go beyond just using the 2012 Olympics to other Olympic games, world championships, and regattas.

Works Cited: See Appendix E

Appendix A: Data Set

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Racer | Home Country | Gender (1=Male 0=Female) | Number of People in the Boat | Heat Number | Semi-Final Number | Event (in meters) | Place in Heat | Place in Semi-final | Place in Final |
| McKeever, Edward | Male | 1 | 1 | 2 | 2 | 200 | 1 | 3 | 1 |
| Rivero, Saul | Spain | 1 | 1 | 3 | 1 | 200 | 3 | 2 | 2 |
| De Jonge, Mark | Canada | 1 | 1 | 1 | 1 | 200 | 5 | 1 | 3 |
| Beaumont, Maxime | France | 1 | 1 | 3 | 2 | 200 | 6 | 4 | 4 |
| Salakhov, Evgeny | Russia | 1 | 1 | 2 | 1 | 200 | 7 | 8 | 5 |
| Dudas, Miklos | Hungary | 1 | 1 | 2 | 2 | 200 | 4 | 5 | 6 |
| Novakovic, Marko | Serbia | 1 | 1 | 2 | 1 | 200 | 2 | 6 | 7 |
| Rauhe, Ronald | Germany | 1 | 1 | 3 | 2 | 200 | 17 | 7 | 8 |
| Larsen, Eirik | Norway | 1 | 1 | 3 | 1 | 1000 | 6 | 3 | 1 |
| van Koeverden, Adam | Canada | 1 | 1 | 1 | 1 | 1000 | 1 | 1 | 2 |
| Hoff, Max | Germany | 1 | 1 | 3 | 2 | 1000 | 4 | 2 | 3 |
| Gustagsson, Anders | Sweden | 1 | 1 | 2 | 2 | 1000 | 12 | 8 | 4 |
| Poulsen, Rene | Denmark | 1 | 1 | 3 | 1 | 1000 | 5 | 5 | 5 |
| Yurenia, Aleh | Belarus | 1 | 1 | 2 | 2 | 1000 | 16 | 3 | 6 |
| Sanchez, Francisco | Spain | 1 | 1 | 2 | 2 | 1000 | 18 | 9 | 7 |
| Brabants,Tim | Great Britain | 1 | 1 | 1 | 1 | 1000 | 8 | 7 | 8 |
| Carrington, Lisa | New Zealand | 0 | 1 | 1 | 3 | 200 | 3 | 1 | 1 |
| Osypenko-Radomska, Inna | Ukraine | 0 | 1 | 3 | 3 | 200 | 5 | 3 | 2 |
| Douchev-Janics, Natasa | Hungary | 0 | 1 | 1 | 1 | 200 | 1 | 2 | 3 |
| Rivas, Teresa | Spain | 0 | 1 | 2 | 3 | 200 | 2 | 6 | 4 |
| Walczykiewicz, Marta | Poland | 0 | 1 | 2 | 1 | 200 | 7 | 4 | 5 |
| Lobova, Natalia | Russia | 0 | 1 | 1 | 2 | 200 | 9 | 5 | 6 |
| Walker, Jessica | Great Britain | 0 | 1 | 2 | 2 | 200 | 6 | 8 | 7 |
| Portela, Teresa | Portugal | 0 | 1 | 3 | 1 | 200 | 13 | 7 | 8 |
| Kozak, Danuta | Hungary | 0 | 1 | 2 | 2 | 500 | 7 | 2 | 1 |
| Osypenko-Radomska, Inna | Ukraine | 0 | 1 | 1 | 3 | 500 | 4 | 4 | 2 |
| Hartley, Bridgitte | South Africa | 0 | 1 | 1 | 2 | 500 | 8 | 3 | 3 |
| Paldanius, Sofia | Sweden | 0 | 1 | 2 | 3 | 500 | 1 | 7 | 4 |
| Idem, Josefa | Italy | 0 | 1 | 3 | 2 | 500 | 14 | 1 | 5 |
| Cawthorn, Rachel | Great Britain | 0 | 1 | 3 | 4 | 500 | 10 | 10 | 6 |
| Hansen, Henriette | Denmark | 0 | 1 | 2 | 1 | 500 | 6 | 6 | 7 |
| Rikala, Anne | Finland | 0 | 1 | 1 | 1 | 500 | 5 | 5 | 8 |
| Postrigay/Dyachenko | Russia | 1 | 2 | 1 | 1 | 200 | 8 | 1 | 1 |
| Piatrushenka/Makhneu | Belarus | 1 | 2 | 2 | 2 | 200 | 1 | 2 | 2 |
| Heath/Schofield | Great Britain | 1 | 2 | 2 | 1 | 200 | 2 | 5 | 3 |
| Hybois/Jouve | France | 1 | 2 | 1 | 2 | 200 | 10 | 4 | 4 |
| Correa/Rezola | Argentina | 1 | 2 | 2 | 1 | 200 | 3 | 6 | 5 |
| Phillips/Bird | Australia | 1 | 2 | 2 | 2 | 200 | 4 | 8 | 6 |
| Cochrane/Fournel | Canada | 1 | 2 | 1 | 1 | 200 | 11 | 7 | 7 |
| Rauhe/Ems | Germany | 1 | 2 | 1 | 2 | 200 | 9 | 3 | 8 |
| Weber/Dietze | Germany | 0 | 2 | 2 | 1 | 500 | 3 | 1 | 1 |
| Krovacs/Douchev-Janics | Hungary | 0 | 2 | 3 | 1 | 500 | 2 | 2 | 2 |
| Naja/Mikolajczyk | Poland | 0 | 2 | 3 | 1 | 500 | 17 | 4 | 3 |
| Wu/Zhou | China | 0 | 2 | 2 | 2 | 500 | 1 | 3 | 4 |
| Schuring/Schwarz | Austria | 0 | 2 | 3 | 2 | 500 | 12 | 5 | 5 |
| Vasconcelos/Gomes | Portugal | 0 | 2 | 3 | 2 | 500 | 7 | 8 | 6 |
| Carrington/Taylor | New Zealand | 0 | 2 | 3 | 1 | 500 | 8 | 6 | 7 |
| Moldovan/Moldovan | Serbia | 0 | 2 | 2 | 2 | 500 | 4 | 9 | 8 |
| Dombi/Kokeny | Hungary | 1 | 2 | 2 | 0 | 1000 | 1 | 0 | 1 |
| Pimenta/Silva | Portugal | 1 | 2 | 2 | 2 | 1000 | 2 | 5 | 2 |
| Hollstein/Ihle | Germany | 1 | 2 | 1 | 0 | 1000 | 3 | 0 | 3 |
| Smith/Wallace | Australia | 1 | 2 | 1 | 1 | 1000 | 7 | 4 | 4 |
| Oscarsson/Nilsson | Sweden | 1 | 2 | 2 | 1 | 1000 | 4 | 3 | 5 |
| Medvedev/Ryakhov | Russia | 1 | 2 | 2 | 2 | 1000 | 9 | 2 | 6 |
| Fitzgerald/Ferguson | New Zealand | 1 | 2 | 2 | 1 | 1000 | 5 | 6 | 7 |
| Gelle/Vlcek | Slovakia | 1 | 2 | 1 | 2 | 1000 | 8 | 1 | 8 |
| Szabo/Kozak/Kovacs/Fazekas | Hungary | 0 | 4 | 1 | 0 | 500 | 5 | 0 | 1 |
| Leonhardt/Weber/Wagner-Augustin/Dietze | Germany | 0 | 4 | 2 | 0 | 500 | 1 | 0 | 2 |
| Pamialova/Papok/Khudzenka/pautaran | Belarus | 0 | 4 | 2 | 0 | 500 | 3 | 0 | 3 |
| walczykiewicz/Konieczna/Naja/Mikolajczyk | Poland | 0 | 4 | 2 | 1 | 500 | 6 | 1 | 4 |
| Walker/Cawthorn/Hannah/Sawers | Great Britain | 0 | 4 | 1 | 1 | 500 | 7 | 4 | 5 |
| Portela/Vasconcelos/Gomes/Rodrigues | Portugal | 0 | 4 | 2 | 0 | 500 | 2 | 0 | 6 |
| Salakhova/Sobetova/Podolskaya/Kachalova | Russia | 0 | 4 | 2 | 1 | 500 | 4 | 3 | 7 |
| Delattre-Demory/Mayer/Guyot/Tuleu | France | 0 | 4 | 1 | 1 | 500 | 8 | 5 | 8 |
| Smith/Smith/Stewart/Clear | Australia | 1 | 4 | 2 | 1 | 1000 | 5 | 1 | 1 |
| Kammerer/Toth/Kulifai/Pauman | Hungary | 1 | 4 | 2 | 0 | 1000 | 1 | 0 | 2 |
| Havel/Trefil/Dostal/Sterba | Czech Republic | 1 | 4 | 2 | 1 | 1000 | 2 | 3 | 3 |
| Gross/Broeckl/Wieskoetter/Hoff | Germany | 1 | 4 | 1 | 1 | 1000 | 4 | 2 | 4 |
| Wraae/Poulsen/Staer/Bleibach | Denmark | 1 | 4 | 1 | 1 | 1000 | 7 | 6 | 5 |
| Gelle/Vlcek/Jankovec/Tarr | Slovakia | 1 | 4 | 1 | 0 | 1000 | 3 | 0 | 6 |
| Medvedev/Ryakhov/Vasilev/Zhestkov | Russia | 1 | 4 | 2 | 1 | 1000 | 6 | 4 | 7 |
| Neagu/Loneticu/Vasile/Gavrila | Romania | 1 | 4 | 1 | 1 | 1000 | 8 | 5 | 8 |

Appendix B: Summary Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Final Finish Place | Semi-Final Place | Heat Place | Semi-Final Number | Heat-Number | Gender | Number in boat | Event in meters |
| Standard Deviation | 2.307367 | 2.646601 | 4.1603 | 0.8334507 | 0.7229742 | 0.50039 | 1.162804 | 333.2989 |
| Mean | 4.5 | 3.841 | 5.958 | 1.403 | 1.889 | 0.5556 | 2 | 566.7 |
| Minimum | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 200 |
| Maximum | 8 | 10 | 18 | 4 | 3 | 1 | 4 | 1000 |

Appendix C: Coding For R

library("xlsx")  
library("car")  
library("bstats")  
#Load in Excel Data  
kayak\_dataset<-read.xlsx("C:/Users/Austin/Documents/Fall 2013 Classes/Applied Stats/kayak data set.xlsx",1)  
#Take Columns in dataset and Make into Variables  
gender <-kayak\_dataset[,3]  
number\_in\_boat <-kayak\_dataset[,4]  
heat\_number <-kayak\_dataset[,5]  
semi\_final <-kayak\_dataset[,6]  
event\_meters <-kayak\_dataset[,7]  
heat\_place <-kayak\_dataset[,8]  
semi\_final\_place <-kayak\_dataset[,9]  
final\_place <-kayak\_dataset[,10]  
summary(gender)  
summary(number\_in\_boat)  
summary(heat\_number)  
summary(semi\_final)  
summary(event\_meters)  
summary(heat\_place)  
summary(semi\_final\_place)  
summary(final\_place)  
sd(gender)  
sd(number\_in\_boat)  
sd(heat\_number)  
sd(semi\_final)  
sd(event\_meters)  
sd(heat\_place)  
sd(semi\_final\_place)  
sd(final\_place)  
#Run Regression 1  
lm(final\_place ~ semi\_final\_place + heat\_place + semi\_final + heat\_number)  
#Run Summary 1  
place\_results <-lm(final\_place ~ semi\_final\_place + heat\_place + semi\_final + heat\_number)  
place\_summary <-summary(place\_results)  
place\_summary  
#Calculate F   
fstat\_place <-place\_summary$fstatistic  
fcrit\_place <-qf(.95,4,67)  
#Hypothesis Test  
hypoth\_rej=0; if(fstat\_place>fcrit\_place){hypoth\_rej=1}  
hypoth\_rej  
#Statistically Signifcant  
####################################################################################  
#Run Regression 2  
lm(final\_place ~ semi\_final\_place + heat\_place + gender)  
#Run Summary 2  
gender\_results <-lm(final\_place ~ semi\_final\_place + heat\_place + gender)  
gender\_summary <-summary(gender\_results)  
gender\_summary  
#Calculate F   
fstat\_gender <-gender\_summary$fstatistic  
fcrit\_gender <-qf(.95,3,68)  
#Hypothesis Test  
hypoth\_rej=0; if(fstat\_gender>fcrit\_gender){hypoth\_rej=1}  
hypoth\_rej  
#Statistically Significant  
#####################################################################################  
#Run Regression 3  
lm(final\_place ~ semi\_final\_place + heat\_place + number\_in\_boat)  
#Run Summary 3  
boat\_size\_results <-lm(final\_place ~ semi\_final\_place + heat\_place + number\_in\_boat)  
boat\_size\_summary <-summary(boat\_size\_results)  
boat\_size\_summary  
#Calculate F   
fstat\_boat\_size <-boat\_size\_summary$fstatistic  
fcrit\_boat\_size <-qf(.95,3,68)  
#Hypothesis Test  
hypoth\_rej=0; if(fstat\_boat\_size>fcrit\_boat\_size){hypoth\_rej=1}  
hypoth\_rej  
#Statistically Significant?  
#####################################################################################  
#Run Regression 4  
lm(final\_place ~semi\_final\_place + heat\_place + event\_meters)  
#Run Summary 4  
meters\_results <-lm(final\_place ~semi\_final\_place + heat\_place + event\_meters)  
meters\_summary <-summary(meters\_results)  
meters\_summary  
#Calculate F   
fstat\_meters <-meters\_summary$fstatistic  
fcrit\_meters <-qf(.95,3,68)  
#Hypothesis Test  
hypoth\_rej=0; if(fstat\_meters>fcrit\_meters){hypoth\_rej=1}  
hypoth\_rej  
#Statistically Significant

Appendix D: Regression Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Statistical Significance Key: | .001 = \*\*\* | .01=\*\* | .05=\* | .01=. | 1= |  |
|  |  |  |  |  |  |  |
| Regression 1 |  |  |  |  |  |  |
|  | Dependent | Intercept | Explanatory 1 | Explanatory 2 | Explanatory 3 | Explanatory 4 |
|  | Final Place |  | Semi-Final Place | Heat Place | Semi-Final Number | Heat Number |
| Coefficient |  | 3.46914 | 0.541 | 0.1562 | -0.6691 | -0.5519 |
| Std Error |  | 0.64431 | 0.09178 | 0.05347 | 0.27694 | 0.29217 |
| Statistically Signicant |  | \*\*\* | \*\*\* | \*\* | \* | . |
|  |  |  |  |  |  |  |
| Regression 2 |  |  |  |  |  |  |
|  | Dependent | Intercept | Explanatory 1 | Explanatory 2 | Explanatory 3 |  |
|  | Final Place |  | Semi-Final Place | Heat Place | Gender |  |
| Coefficient |  | 2.01699 | 0.4257 | 0.13741 | 0.04781 |  |
| Std Error |  | 0.50185 | 0.08918 | 0.05672 | 0.43619 |  |
| Statistically Signicant |  | \*\*\* | \*\*\* | \* |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Regression 3 |  |  |  |  |  |  |
|  | Dependent | Intercept | Explanatory 1 | Explanatory 2 | Explanatory 3 |  |
|  | Final Place |  | Semi-Final Place | Heat Place | Number in Boat |  |
| Coefficient |  | 0.60605 | 0.50377 | 0.14966 | 0.53205 |  |
| Std Error |  | 0.66333 | 0.0891 | 0.5395 | 0.1918 |  |
| Statistically Signicant |  |  | \*\*\* | \*\* | \*\* |  |
|  |  |  |  |  |  |  |
| Regression 4 |  |  |  |  |  |  |
|  | Dependent | Intercept | Explanatory 1 | Explanatory 2 | Explanatory 3 |  |
|  | Final Place |  | Semi-Final Place | Heat Place | Event in Meters |  |
| Coefficient |  | 1.6880749 | 0.4408735 | 0.1326899 | 0.0005738 |  |
| Std Error |  | 0.5996532 | 0.0904707 | 0.0566895 | 0.0006665 |  |
| Statistically Signicant |  | \*\* | \*\*\* | \* |  |  |

Appendix E: Works Cited

Nibali, Maria, Will Hopkins, and Eric Drinkwater. "Variability and Predictability of Elite Competitive Slalom Canoe-kayak Performance." European Joirnal of Sport Science 11.2 (2011): 125-31. Ebsco Host. Web.

Davranche, Karen. "Decision Making in Elite White-Water Athletes Paddling on a Kayak Ergometer." Journal of Sport & Exercise Pyschology 31.4 (2009): 554-66. Ebsco Host. Web

Gomes, B.B. "Gross Efficiency and Energy Expenditure in Kayak Ergometer Exercise." International Journal of Sports Medicine 33.8 (2012): 654-61. Ebsco Host. Web

Ridge, Barry. "Morphological Characteristics of Olympic Slalom Canoe and Kayak Paddlers." Eurpoean Journal of Sport Science 7.2 (2007): 107-14. Ebsco Host. Web.

Bishop, D., D. Bonetti, and B. Dawson. "The Influence of Pacing Strategy on VO2 and Supramaximal Kayak Performance." Medicine & Science in Sports & Exercise 34.6 (2002): 1041-047. Web.

"London 2012 Results Archive." SportScene.