**Analysis of Athletes’ Performance in the Hawaiian Ironman Championships from 2011-2015.**

Aditya Gautam

**Abstract**

Analysis has been conducted on the time taken to finish each event in the Ironman Championships from 2011-2015. The data regarding the athletes which recorded an either, did not start (DNS) or did not finish (DNF) in any event or the whole race altogether. The results of the linear regression models which tell us the most and the least influential events in each year have been illustrated. The effects aging on the performance of the athletes have been discussed. These were analysed using the method of non-linear curve fitting which gave us the parameters of decline and the age when the performance of the athlete equals zero.

*Keywords:* Swim, Bike, Run, Time, Baseline, Division, Performance, Metric, Decline.

**Introduction**

**Background**

The Ironman race was originally conceived in 1977 with swimming and running enthusiasts in Hawaii debating who was more fit, at the same time a Sports Illustrated article lauded a Belgian cyclist for having the highest ‘oxygen uptake’ recorded in history. Thus, a race around Kona was developed *“Swim 2.4 miles, Bike 112 miles and Run 26.2 miles! Brag the rest of your life”* the winner was crowned an **Ironman.** Triathlon is now one of the fastest growing sports for amateurs. The Hawaiian Ironman Championship is the only sports event in the world to include the top professionals and qualifying amateurs racing side by side, any man or woman can win. The data was taken from ironman.com.

**Research Questions**

With the amount of time, money and effort put in by athletes, we attempt to answer the following questions:

* Which segments of the race impact the results of the race most?
* Performance deterioration due to age?
* Has the level of championships increased or decreased over the set period?
* What race segment was the most difficult in each year of the championship?

**Motivation**

The fact that amateurs and professionals compete in the same event gives us very diverse dataset and the results of this study will hopefully, help us in analysing the qualifying events that are held throughout the world. The relation of the performance deterioration with age can assist us in better understanding of the training methods that an amateur athlete has to go through in order to compete in the race. There have been many studies which represent physiological decline and oxygen uptake as the body ages, for professional athletes. But very few studies have been conducted regarding the amateur participants. The results will also help athletes devise an optimal strategy to compete in the race.

**Literature Review**

The main idea for this study came from the ‘Aging Performance Decline in Masters Records in Athletics Swimming, Rowing, Cycling, Triathlon and Weightlifting by A. Barry Baker and Yong Q. Tang’. In this study the Record performances for Masters sporting events were analysed and then a comparison study with the authors’ previously published study for Masters running, walking, and jumping sports events is published. Records were normalized using the 30s age records as a baseline and studied through the various age ranges to the 90s. A curvilinear mathematical model [ ]. All sports performances declined with increasing age with rowing showing the least decline and weightlifting represented the most decline in performance with increasing age. The main difference between our analysis and this study here is that we have analysed the performances of amateur athletes only and the results regarding the professional athletes were omitted.

Hopefully, this study will shed some light on the healthy lifestyle and finding out an optimal exercising routine for the population. In many countries, especially in the western world, there is a growing number of elderly people (>65 years old) each decade and along with this increase in this demographic there is a continuous decrease in the birth rate. These elderly people are exercising more, as there is a correlation in medical care and appreciation of lifestyle that impact and enhance life expectancy. Along with this enhanced longevity there is an ever-increasing interest and motivation in exercising the aspirations of many amateur athletes has turned towards these Ironman races. The events in these races require less skill and more endurance. This could be the main reason that we see the participation numbers as demonstrated in this study and therefore, attempting the World Records in their respective age group. Baker and Tang have emphasised on the fact that the athletes must be extremely motivated, well trained and practiced at the sporting activity they undertake. Similar assumptions must be made regarding the participants of the Ironman races.

This study of these highly motivated and fit individuals gives a nice indication of the best possible performance for the age groups concerned and the decline of performances with age. Using the data for only the ‘Hawaiian Ironman Championships’ removes the inconsistencies which may arise due to different altitude levels, longitudinal time differences, environmental conditions and health factors. But, undertaking of such longitudinal studies might give us a better indication of the averaging age potential rather than the best age potential.

The article by Christopher Minson (Professor at University of Oregon), discusses how the ability of the human body to utilize oxygen changes as age increases. In most sports the ‘sweet spot’ age, where the physical, technical and strategic are at the highest level of synchronisation. In most sports this sweet spot falls in mid-20’s to early-30’s. However, there are many examples of athletes winning gold medals at the Olympics in their late 40’s and 50’s. But, the vast majority of the sporting events in which they were competing required the possession of exceptional skill and less aerobic and anaerobic powers, such as shooting and equestrian. For endurance sports the upper cap for competing at the highest level was around the age of 40. This can’t be said of the Ironman races as it is opened to all the best professionals and all the amateur athletes who earn their spot at the World Championships by going through qualifying events held around the world. This article gives a few examples, where the winners were athletes who were either about to turn 40 or just crossed the 40-year mark.

Discussion further continues regarding the ability of the body to utilise the oxygen effectively. This ability has been named as VO2max which is the predictor of endurance performance ages across ages. VO2max is a numerical value that describes how much oxygen your body can use per kilogram of body weight. It is affected by how well your body can bring oxygen in the lungs and how well it is carried to the muscles by our blood. In general population the VO2max declines by about 10% per decade after the age of 30. This decline can be halved if the person continues to train rigorously. The main reason of this decline is the maximal heartrate, which also decreases as we age. This means that the blood is not carrying enough oxygen in the body as the age increases. Even if oxygen delivery goes down, our muscles adapt appropriately to utilize the oxygen effectively relative to the given workload. This is well maintained into our 60’s and 70’s. It explains why there are so many participants above the age of 60 at the Ironman races throughout the world. The findings of this article prompted the age group selection for the study of performance decline.

From this article it is safe to make a claim, that athletes or any normal human being should train ‘smarter’ not ‘harder’. A perfectly designed training regime will help athletes recover better and reduce their chances of getting injured. However, this goes without mentioning that, the training method needs to evolve as age increases.

In Joseph Baker’s study ‘Do or Decline’ he has resonated the facts from the previous articles. Baker points to a seminal 1996 study from Stanford University analysing age-related decline that looked at areas such as the number of muscle cells, DNA repair, fingernail growth and physical activity. The finding was that there is a 0.5 percent decline per year, a statistic he says has served as the biomarker of the aging process. The idea of researching data from sports events is the best way to study the effects of aging on the human body.

**Data and Methodology**

The observations regarding the athletes which either failed to finish an event or the full race (DNF) or they did not start an event (DNS) have been removed from the datasets. So, for each year’s championship from 2011 to 2015 we have the data of the athletes who finished the race. Here is the brief description of the data:

|  |  |
| --- | --- |
| **Championship Year** | **Dataset name in the notebook** |
| 2011 | race1 |
| 2012 | race2 |
| 2013 | race3 |
| 2014 | race4 |
| 2015 | race5 |

*Table 1: datasets.*

For computational purposes all the event times were converted into hours (e.g. if the time taken to complete the 26.2-mile run is 03:00:50 it is represented as 3.0139 hrs)

**Descriptive Statistics**

For 2011, total participants were 1918 out of which 1774 reached the finishing line i.e. 92.50% of the total.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **age** | **swim\_time** | **bike\_time** | **run\_time** | **overall\_time** |
| **mean** | 46.47682 | 1.238938 | 5.707424 | 4.135234 | 11.24073 |
| **std** | 10.77174 | 0.238178 | 0.748287 | 0.906719 | 1.76235 |
| **min** | 16 | 0.83 | 4.31 | 0.03 | 8.07 |
| **25%** | 39 | 1.07 | 5.17 | 3.51 | 9.98 |
| **50%** | 45 | 1.19 | 5.53 | 3.89 | 10.78 |
| **75%** | 53 | 1.35 | 6.03 | 4.47 | 11.9275 |
| **max** | 86 | 2.29 | 8.7 | 7.92 | 17 |

*Table2: Descriptive Statistics for race1 dataset*

For 2012, out of 2038, 1887 participants crossed the finish line i.e. 92.60% of the total.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **age** | **swim\_time** | **bike\_time** | **run\_time** | **overall\_time** |
| **mean** | 46.32279 | 1.244035 | 5.897413 | 4.239104 | 11.54138 |
| **std** | 11.34955 | 0.242213 | 0.759769 | 0.933652 | 1.804528 |
| **min** | 22 | 0.1389 | 4.4303 | 2.2786 | 8.3103 |
| **25%** | 38 | 1.0736 | 5.31765 | 3.5768 | 10.2339 |
| **50%** | 45 | 1.18515 | 5.68085 | 3.9653 | 11.0333 |
| **75%** | 53 | 1.361975 | 6.35865 | 4.6493 | 12.40205 |
| **max** | 86 | 2.4489 | 8.7442 | 8.5336 | 16.9886 |

*Table3: Descriptive Statistics for race2 dataset*

For 2013, out of 2134 participants 1973 crossed the finish line i.e. 92.45% of the total

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **age** | **swim\_time** | **bike\_time** | **run\_time** | **overall\_time** |
| **mean** | 45.10497 | 1.200276 | 5.610132 | 4.153882 | 11.10619 |
| **std** | 11.06745 | 0.219607 | 0.757288 | 0.914798 | 1.741499 |
| **min** | 3 | 0.8272 | 4.3639 | 2.14 | 8.2081 |
| **25%** | 37 | 1.0518 | 5.052175 | 3.52 | 9.8417 |
| **50%** | 44 | 1.1531 | 5.39375 | 3.88 | 10.6194 |
| **75%** | 52 | 1.29905 | 5.977225 | 4.52 | 11.8378 |
| **max** | 83 | 2.4378 | 9.3642 | 8.31 | 16.9481 |

*Table4: Descriptive Statistics for race3 dataset.*

For 2014, out of 2187 participants 1985 crossed the finish line i.e. 90.76% of the total.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **age** | **swim\_time** | **bike\_time** | **run\_time** | **overall\_time** |
| **mean** | 44.8344 | 1.26491 | 6.022575 | 4.168972 | 11.59367 |
| **std** | 11.40745 | 0.254254 | 0.903202 | 0.863957 | 1.868554 |
| **min** | 19 | 0.85 | 4.35 | 2.41 | 8.24 |
| **25%** | 37 | 1.09 | 5.29 | 3.55 | 10.18 |
| **50%** | 44 | 1.205 | 5.8 | 3.93 | 11.15 |
| **75%** | 52 | 1.39 | 6.5775 | 4.6 | 12.49 |
| **max** | 86 | 2.48 | 9.03 | 7.58 | 16.94 |

*Table5: Descriptive Statistics for race4 dataset.*

For 2015, out of 2308 participants 2143 crossed the finish line i.e. 92.85% of the total.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **age** | **swim\_time** | **bike\_time** | **run\_time** | **overall\_time** |
| **mean** | 43.58231 | 1.278611 | 5.899121 | 4.373929 | 11.70579 |
| **std** | 11.18392 | 0.2485 | 0.826615 | 0.912766 | 1.806772 |
| **min** | 20 | 0.85 | 4.42 | 2.41 | 8.24 |
| **25%** | 35 | 1.1075 | 5.25 | 3.71 | 10.35 |
| **50%** | 43 | 1.23 | 5.72 | 4.11 | 11.25 |
| **75%** | 51 | 1.41 | 6.43 | 4.81 | 12.7 |
| **max** | 86 | 2.54 | 8.83 | 8.73 | 16.82 |

*Table5: Descriptive Statistics for race5 dataset.*

**Analysis**

Linear Regression *(y = ax + b)* was used to analyse the data further. Where:

* *y:* overall\_time.
* *x:* swim\_time, bike\_time, run\_time and age. *Lepers R (2008)*
* Formula in the code ('overall\_time~ bike\_time + swim\_time + run\_time + age')

The R-sq. values for the regression models are as follows:

|  |  |
| --- | --- |
| **2011** race1 | 0.999 |
| **2012** race2 | 0.999 |
| **2013** race3 | 0.999 |
| **2014** race4 | 0.959 |
| **2015** race5 | 0.999 |

*Table6: R-squared values.*

The coefficient of age variable showed some interesting characteristics, the lowest value observed was -8.76e-05 with p-value equalling 0.539 for the year 2011 and the highest being 0.0005 with p-value equalling 0.01 for the year 2015. So far from the compiled results it is evident that the age variable has very less influence on the models as its coefficients are very low and the varying p-values question its selection as the independent variable. This also justifies the need for an alternative method to find the performance decline relative to the age. However, the most difficult event can be identified by comparing coefficients associated with each event in the linear regression results. Similarly, the most and least influential events for each year can also be found out. They are as follows:

|  |  |  |
| --- | --- | --- |
| **Event** | **Most difficult** | **Least Difficult** |
| **Biking** | 2011 | 2014 |
| **Swimming** | 2012 | 2014 |
| **Running** | 2014 | 2012 |

*Table7: Race difficulty*

|  |  |  |
| --- | --- | --- |
| **Year** | **Most influential event** | **Least influential event** |
| **2011** | Swimming (1.0447) | Running (1.0312) |
| **2012** | Swimming (1.0572) | Running (1.0214) |
| **2013** | Swimming (1.0415) | Biking (1.0297) |
| **2014** | Running (2.29) | Biking (0.2538) |
| **2015** | Swimming (1.0509) | Biking (1.0316) |

*Table8: Events in each year along with their coefficients in the regression models.*

**Non-Linear Curve Fitting**

The data recorded was analysed to determine the percentage decline in the maximum physiological performance with increasing age. Data was normalized using the age 30-35 records as the baseline and studied through the various age ranges to the maximum which was 86. The observations regarding the ‘PRO’ division were removed as their performances are not comparable with the amateur participants. The selection of age range of 30-35 is because it is usually the time of life when the athletes have passed their peak physical conditions. Only top 10 performers were considered for baselining as this age range appears throughout the dataset from top to bottom in all the championship years. The time data regarding the events swimming (2.4 miles), biking (112 miles) and running (26.2 miles) were analysed.

Deterioration in performance with age is assessed by fitting a non-linear mathematical model of the form where:

* *y:* functional performance
* *T:* age
* *T0:* age when performance equals to fraction of zero.
* *τ:* time constant of curvilinear decline

**Baselining**

Performance data for timed events were calculated as the reciprocal of time. For example, if the time was 25s for 30-35 age range, but 25.5s for 35-40 age range the functional performance for the 35-40 age range would be: (1/25.5)/(1/25) = 0.9803 of the baseline performance. *(Baker et al., 2003)*.

**Results**

First the analysis was conducted for each of the years for each event in the Ironman triathlon championships. The following curves resulted from 2012 data.

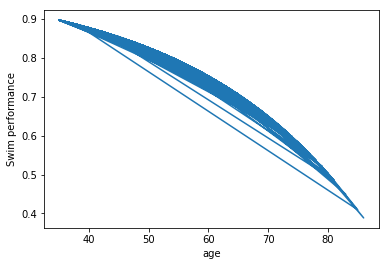


Figure 1 Performance decline with age in swimming for 2012

As predicted the fitted curve gave the values of *T0= 100.11* (the age where the performance equals to fraction of zero) *and τ= 28.69* (the time constant of curvilinear decline) for the swimming event*.* Similarly, this method was applied on each event which gave us the whole range of observations regarding the T0 and τ. The decline along with the age where the performance equals fraction of zero is different for different events.

For 2012 biking event, the values of the parameters were *T0= 105.62 and τ= 27.43.* For the running event the parameters’ values were *T0=100.30 and τ= 36.93*. The curves for each of these events present the decline appropriately.

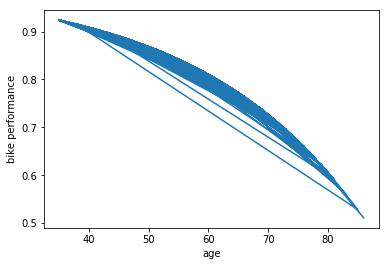


Figure 2: Performance decline with age in biking for 2012.

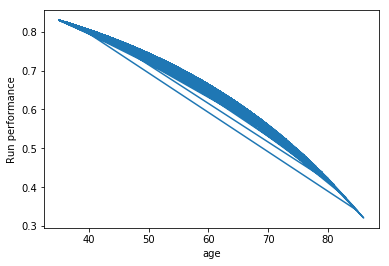


Figure 3: Performance decline with age in running for 2012

The results so far indicate that biking has demonstrated a decline which is only fractionally different from the other two events.

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Event** | **T0** | **τ** |
| 2011 | Swimming | 98.49 | 25.34 |
|  | Biking | 109.92 | 30.59 |
|  | Running | 104.49 | 36.68 |
| 2012 | Swimming | 100.11 | 28.69 |
|  | Biking | 105.62 | 27.44 |
|  | Running | 100.30 | 36.93 |
| 2013 | Swimming | 99.00 | 28.32 |
|  | Biking | 105.75 | 30.22 |
|  | Running | 97.53 | 32.87 |
| 2014 | Swimming | 95.14 | 26.77 |
|  | Biking | 101.17 | 29.58 |
|  | Running | 99.00 | 35.32 |
| 2015 | Swimming | 97.48 | 29.02 |
|  | Biking | 106.23 | 32.83 |
|  | Running | 99.27 | 37.46 |

*Table9: Performance decline parameters of the curve.*

The ‘τ’ is in the range of 25.34 to 37.46 across all events in the years 2011-2015. Whereas, the ‘T0’ is in the range of 95.14 to 109.92. The event which showed the highest decline was ‘Running’ in 2015 and the event which showed the least decline was ‘Swimming’ in 2011. The maximum age where the performance decline equalled 100% was 109.92 for the event ‘Biking’ in 2011. Whereas, the minimum age where the performance decline equalled 100% was 95.14 for ‘Swimming’ in 2014.

There is clear evidence that the amateur athletes hit the 100% decline mark latest in biking event. However, the evidence regarding swimming and running events are comparable and there is no clear method, for this data, to clearly make a decision regarding which of the two events is more difficult. The age range for biking event is the widest as compared to the other two events.

To get a better understanding of performance decline with age for this duration of five years, the data from each year was divided in to age groups 30-39, 40-49, 50-59, 60-69, 70-79 and 79-86. Top 50 performers were selected from each of the group for equal representation purposes. The data was baselined using the 30-39 age data. The results are as follows:

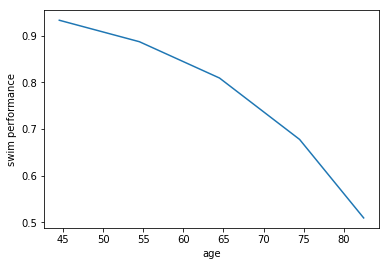


Figure 4: Performance decline from 2011 to 2015 in swimming

T0 = 96.06 and τ = 19.03

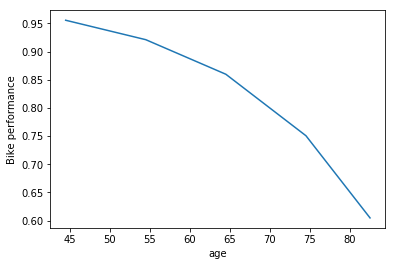


Figure 5: Performance decline from 2011 to 2015 in biking

T0 = 98.62 τ = 17.37

These results show that the decline in all the three events is consistent and comparable. But, it should be kept in mind that the data here was re-configured to for computational purposes.

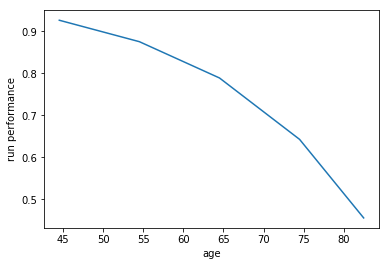


Figure 6:Performance decline from 2011 to 2015 in Running

T0 = 94.07 τ = 19.00

**Conclusion**

The reasons regarding, why a particular event was more difficult or influential in predicting the overall time remains to be seen. This could be related to the environmental conditions, weather conditions, or some other factors. The results illustrated here may prompt a study, how these factors affect the result of the Ironman or any outdoors athletics event. Swimming was the most difficult event in 4 out of the 5 years, more data such as swim velocity of participants, velocity of the current and direction of the swim will shed some light on why this was the most difficult event. Same can be said regarding biking and running if the velocity data for both these two events is also collected, the constant of decline (τ) can be better explained and give us some insight on the physiological decline with age.

The baseline age for this normalization procedure was again taken as the 30s age performance level for the same reason as given in the literature review before. There was in general a greater an abrupt decline in performance between the younger age groups and the 30s age performance, than for the 30s performance versus later age performances. Same was the reason why the data regarding the professional division was not considered for the performance decline study.

A gender-based performance study can be done using this data if they could mention it in the future events. The difference in performance decline and the physiological decline such as, slowed reaction time, decreased coordination and joint mobility, decreased skeletal size and muscle bulk will certainly give a better insight on the body condition of the participants.

**References**

Baker AB, Tang YQ (2010) *Aging performance for masters records in athletics, swimming, rowing, cycling, triathlon, and weightlifting. Exp Aging Res 36:453–77*

Romuald Lepers (2008) *Analysis of Hawaii Ironman Performances in Elite Triathletes from 1981 to 2007 INSERM U887, Faculty of Sport Sciences, University of Burgundy, Dijon, FRANCE*

Christopher Minson (2016) *How does aging affect athletic performance?*

[*https://theconversation.com/how-does-aging-affect-athletic-performance-36051*](https://theconversation.com/how-does-aging-affect-athletic-performance-36051)

Joseph Baker (2010) *Do or Decline.  Journal of Health Psychology*

[*https://www.washingtonpost.com/news/inspired-life/wp/2017/02/03/do-or-decline-an-athletes-age-may-be-less-important-to-performance-than-persistent-practice/?noredirect=on&utm\_term=.492adc2719d2*](https://www.washingtonpost.com/news/inspired-life/wp/2017/02/03/do-or-decline-an-athletes-age-may-be-less-important-to-performance-than-persistent-practice/?noredirect=on&utm_term=.492adc2719d2)