Come One, Come All: The 2016 Super Tread Olympic Triathlon

Summary

Every race begins with a large group of people trying to start from one point. The farther each contestant goes, the more they separate from other contestants. Therefore, we decided to focus our efforts on minimizing the maximum flow rate for the swimming leg of the race and the first bottleneck, transition stop one.

To optimize the flow rate, we first obtained means, standard deviations, and skews. We obtained values for skew using Pearson's Skewness Test. If the skew obtained from this test is between -2 and 2, it is not statistically significant. We graphed the normal distributions of waves using normal variance because no values during the first leg of the race and T1 were significantly skewed. Furthermore, we extended this normal variance to each separate wave because each wave had more than 30 athletes ($n > 30$). We used the probability density of each wave's normal distribution to model the times that athletes from separate waves reached the same points in the race.

We decided to organize the athletes in order of descending average final time, meaning all professionals and premiers were combined in the first wave whereas Athenas and Clydesdales were combined into the final wave. Using the full maximum time allotted, we found that each wave should be separated by 40 seconds. The exception to this standard is female opens, who should be separated by 53 seconds because they have the smallest standard deviation meaning they have the least variance in swim times. Although our model works well in this scenario, the model does not take into account the fact that females have faster swim times than males. The final wave of 50 male opens could be temporarily surpassed by the first wave of 50 female opens during the swimming leg. However, after T2 the males would overtake the females once again, causing race congestion.

Cover Letter

Dear Mayor Inglehoff,

Thank you once again for allowing the YourTown youth organization to host the 2016 Super Tread Olympic Triathlon. We immensely appreciate the copious amounts of land, road, and time you have graciously lent us to run this event. We plan on investing 100% of the proceeds this triathlon accrues in the youth of YourTown, and we hope that the money raised in this event will go a long way towards guiding them into becoming the leaders of tomorrow.

We are confident that this event will be an enormous success because we focused on mitigating race congestion: we decided to make sure that no one checkpoint will bottleneck the triathlon and as a result have multiple competitors waiting on competitors of a previous wave to proceed. A pleasurable experience is key to the success of this triathlon. If a professional is in any way inhibited by open male or female contestants, the triathlon will not be enjoyable for fans or our sponsor.

Nevertheless, while prioritizing race congestion we also took into account the entire time period in which the triathlon will take place. Once again, we are very thankful for the roads you have let us use, and we fully realize the disturbance in the normal routine of the town shutting down roads will temporarily cause. We designed the wave timings (see page 2) to return the town to normalcy as soon as possible without sacrificing the success of the event itself. To do this, we first calculated the means and standard deviations of the total times and times at each stage of the triathlon for each skill level. By testing for skew, we found that the times for each category of racer, the final times, as well as times for each leg of the triathlon, were randomly distributed around the average according to a statistical model known as the normal distribution. We then created a model of the flow rate of people through the entrance to the first transition for each wave using the normal distribution. By adding all these models together, we produced an algorithm for the maximum flow rate for any given wave size and wave spacing.

By adjusting the parameters of wave size and wave spacing, we reduced the flow through what we believe is the largest bottleneck, the end of the swim leg, to 1.2 people per second. Using the wave schedule determined by our model, the triathletes will be on and off the roads in your town within 5.5 hours. In the future if we could choose the length of our triathlon, we would increase the swimming distance and biking distance and decrease the running distance. This would reduce the flow rate through the transition points, and also reduce the time that the roads would need to be closed. Thank you for your support of this triathlon, and I hope we can work with each other to provide for our youth with more triathlons in future years.

Regards,
Team 6284

Introduction

In the city named "YourTown," a triathlon is held to raise money for the local youth organization. Fortunately, the triathlon has been fully funded by Super Tread Race Company and the mayor has given us permission to proceed. However, both have certain preferences and requirements for how the triathlon should be run. Super Tread wants the triathlon to attract pros and premiers and become a world class event. Race congestion must be minimized to please enough pros and premiers for Super Tread to fund next year. The time in which the roads are closed is also a priority. Although Mayor Inglehoff wants the event to be a success, he has asked us to keep the road closure time below 5.5 hrs. The mayor and Super Tread have asked us to draft a heat schedule which reduces race congestion and road closure time. We will also

create divisions to break up the participants for scoring purposes. Further, we will try to adjust the lengths of the legs of the triathlon and analyze the consequent effects of race congestion and road closure time.
**In our model all athletes must cross the finish line within 5.5 hours of the first athlete entering the biking section.
The flow rate of athletes through every point must be minimized, and the frequency of athletes passing each other must be minimized.

Assumptions

A few assumptions were made regarding the triathlon to help better quantify the problem.

- The given data set accurately represents the entire population of triathlon athletes.

  Justification: The given data set is of 3217 athletes at a previous triathlon not hosted by YourTown. Our triathlon is projected to have 2000 athletes. In order to best quantify the problem, an assumption must be made that all possible data of any athlete is all represented in just the 3217 athletes in the previous triathlon.

  - Athletes participating in the triathlon are 18 or older.

    Justification: The given data set has all athletes of 18 years or older. Based on the first assumption, there is no record of athletes younger than 18, hence they cannot join the triathlon.

- Roads remain open until the first athlete leaves T1.

  Justification: The first portion of the triathlon is 1500 m swim. Roads will not be in use during the swimming portion of the triathlon. The swimming portion starts when the first athlete enters the water and continues until the first athlete begins biking. During this time, roads may be used by YourTown residents.

- No major delays, failures, or malfunctions affect the duration of the race.

  Justification: Unfortunately, we cannot accurately predict when, where, and what major accident will happen. The triathlon will have emergency services on standby if something unexpected happens. These services will act immediately and will add no time to the duration of the triathlon.

- Waves are not limited to the divisions

  Justification: Divisions are based on the age range and category the athlete is in. However, these divisions alone cannot form effective waves. So, multiple divisions are grouped into one wave and multiple waves have the same division multiple times.

- On the road, the athletes never visit a point previously visited.

  Justification: The triathlon athletes only travel in one direction, and never turn around. If they did, it would unnecessarily double race congestion because pros would be either running or biking into open participants.

Model Description

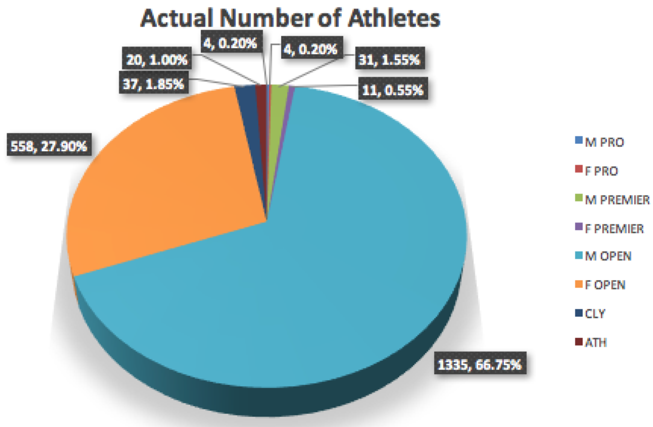


Figure 1: *The Number of Participants in the Triathlon* The pie chart above shows the total number of participants in the 2016 Super Tread Olympic Triathlon divided into their categories. The projected number of participants is 2000.

Table 1: Divisions

Divisions	Number of Particpants
Male Pros	4
Female Pros	4
Male Premiers	31
Female Premiers	11
Male Open 18-29	175
Male Open 30-39	465
Male Open 40-49	397
Male Open 50-59	233
Male Open 60-69	58
Male Open 70-79	7
Female Open 18-29	98
Female Open 30-39	220
Female Open 40-49	155
Female Open 50-59	70
Female Open 60-69	14
Female Open 70-79	1
Athena	37
Clydesdale	20

Table 2: Mean for each segment of the triathlon for each skill level

AVG	M	F	CLY	ATH
Avg Swim Pro	0:13:08	0:12:58	n/a	n/a
Avg T1 Pro	0:03:16	0:03:38	n/a	n/a
Avg Final Pro	1:54:46	2:03:28	n/a	n/a
Avg Swim Premier	0:16:40	0:17:02	n/a	n/a
Avg T1 Premier	0:04:24	0:04:57	n/a	n/a
Avg Final Premier	2:10:42	2:23:49	n/a	n/a
Avg Swim Open	0:23:21	0:22:29	0:25:38	0:23:22
Avg T1 Open	0:07:38	0:10:02	0:08:03	0:11:37
Avg Final Open	2:58:12	3:16:49	3:16:45	3:45:58

Table 3: Standard Deviation for each segment of the triathlon for each skill level

	Final	Swim	T1
Male Pros	0:05:55	0:01:13	0:00:20
Female Pros	0:04:27	0:00:21	0:00:10
Male Premiers	0:08:38	0:01:26	0:00:34
Female Premiers	0:08:24	0:01:24	0:00:38
Male Open	0:26:26	0:04:16	0:03:00
Female Open	0:30:36	0:03:21	0:03:26
Athenas	0:38:04	0:03:33	0:03:39
Clydesdales	0:30:03	0:03:50	0:03:12

Other Solutions

Additional Assumptions

- For the purpose of these solutions, each road individually can remain closed for only 5.5 hours, but the total time in which roads remain closed can exceed 5.5 hours.

Solution 2

In addition to the main solution, other methods to obtain potential solutions were tried as well. In this solution, the group with the highest average time (see Table 1), the Athenas, went first and the successive waves went in descending average time all the way to the male pros. Unfortunately, this mathematical model did not work out as well as expected. First, the maximum time for wave 1 (which was Athenas) to three standard deviations and the minimum time for wave 2 (which was open females) to three standard deviations were calculated (see Table 2). We used the following equations to obtain max and min values. Then, the difference between the max Athena time and the minimum female open time was calculated.

$$MaxAthenaTime = MeanAthenaTime + 3 * StdevAthenaTime$$

$$MaxAthenaTime = 03:45:58 + 3 * 00:38:04$$

$$MaxAthenaTime = 05:40:10$$

$$MinFemaleOpenTime = MeanFemaleOpenTime - 3 * StdevFemaleOpenTime$$

$$MinFemaleOpenTime = 03:16:49 - 3 * 00:30:36$$

$$MinFemaleOpenTime = 01:45:01$$

$$TimeBetweenHeats = MaxAthenaTime - MinFemaleOpenTime$$

$$TimeBetweenHeats = 05:40:10 - 01:45:01$$

$$TimeBetweenHeats = 3:55:10$$

This difference between the maximum Athena time and the minimum female open time is the time necessary to wait between wave one and wave two in order to make sure that no female open passes or gets inhibited by the slowest Athena. As shown above, the time necessary to wait just between wave one and wave two is 3 hours 55 minutes and 10 seconds. Because there are going to be over 20 waves (max 100 people waves for an estimated 2000 participants), waiting to unleash the second wave this long doesn't make any sense. In addition, the maximum open female time to three standard deviations is 04:48:37, meaning she would cross the finish line at 08:43:46, rendering this solution invalid.

Solution 3

On the contrary, our second alternate solution worked in under the maximum time allotted to us by the mayor, given the extra assumption. To start, the maximum size of a wave was set as 100. Next, the number of waves needed was calculated from the number of each type of participant. The actual sizes were scaled down from the original data considering there were 3217 participants in the given data, but only ~2000 participants are expected at The 2016 Super Tread Olympic Triathlon.

The first wave of this solution started with male pros, then alternated gender going down waves. To calculate the time between waves, the same method as before was used for each step, and the wave schedule below was formed. Each wave where the division changes (e.g. Male Pro to Female Pro) is released a few seconds after the difference between the maximum from the first wave and the minimum of the second wave to maintain aesthetically pleasing times.

Table 5: The wave schedule for the event along with the minimum difference required between different groups to ensure they don't collide and cause race congestion.

WAVES	GROUP	ESTIMATED SIZE	TIME	Min diff btw waves
WAVE1	M PRO	5	00:00:00	00:02:03
WAVE2	F PRO/M PREMIER	35	00:02:05	0:08:09
WAVE3	F PREMIER/ M OPEN	100	0:10:10	0:00:30
WAVE4	M OPEN	100	0:10:40	0:00:30
WAVE5	M OPEN	100	0:11:10	0:00:30
WAVE6	M OPEN	100	0:11:40	0:00:30
WAVE7	MOPEN	100	0:12:10	0:00:30
WAVE8	M OPEN	100	0:12:40	0:00:30
WAVE9	M OPEN	100	0:13:10	0:00:30
WAVE10	M OPEN	100	0:13:40	0:00:30
WAVE11	M OPEN	100	0:14:10	0:00:30
WAVE12	M OPEN	100	0:14:40	0:00:30
WAVE13	M OPEN	100	0:15:10	0:00:30
WAVE14	M OPEN	100	0:15:40	0:00:30
WAVE15	M OPEN	100	0:16:10	0:00:30
WAVE16	M OPEN/CLY	87	0:16:40	0:14:34
WAVE17	F OPEN	100	0:31:15	0:00:30
WAVE18	F OPEN	100	0:31:45	0:00:30
WAVE19	F OPEN	100	0:32:15	0:00:30
WAVE20	F OPEN	100	0:32:45	0:00:30
WAVE21	F OPEN	100	0:33:15	0:00:30
WAVE22	F OPEN/ ATH	79	0:33:45	

The largest amount of time an Athena would be on a road (not including her time in the water or at T1) to three standard deviations was calculated to be 04:46:54. This plus the starting time for the Athena, 0:33:45, yields a finishing time of 5:17:16, which falls under the 5:30:00 maximum time.

Table 4: Skew for each segment of the triathlon for each skill level

	M PRO	F PRO	M PREMIER	F PREMIER	M OPEN	F OPEN	ATH	CLY
Average Final Time	0.254	-0.162	0.494	0.188	0.922	1.063	0.367	0.371
Swim Time	1.211	0.82	0.366	0.726	1.228	1.138	0.589	0.402
T1	1.092	1.339	1.195	0.692	1.677	1.181	1.122	1.339

Optimal

In order to reduce congestion, we need to release waves in a way that reduces the maximum flow rate, in people per second, throughout the entire race. First we found the number of contestants in each category, the mean, median, minimum, maximum, and standard deviation, of the times for each category of athlete for each leg. From this we only used the mean, standard deviation, and skew of the data, and considered the min and max of the given triathlon to be unrepresentative of our triathlon.

The data for each category of athlete besides professionals are biological data, which we expected to be distributed along the normal distribution. We used the Phearson's Skewdness Test to determine the degree to which the data was normally distributed, testing each leg of the triathlon for each category and taking significant skew to be s<2. From this we found that only the times for the Clydesdales were significantly skewed. Because we have a large sample size for the open categories and heavyweight categories, and the data aren't significantly skewed, we considered the data to be normally distributed. Therefore, we can consider waves of more than 30 athletes to be representative of the total population and therefore also normally distributed.

If the athletes are normally distributed, then we can model the times that they will finish any given leg of the triathlon by the probability density function for the distribution, namely

$$\frac{s_2}{\sqrt{2\pi\sigma_2{}^2}} e^{\frac{(t-d)^2}{2s_2{}^2}}$$

S=number of people in the wave
Sigma=standard deviation of time from start to point
d=delay from beginning of race to start of wave
By adding the probability densities of each wave, we calculated the total probability of a person arriving at the point at any time, which represents the expected total flow rate at that point and time.

$$p = \sum_{n=0}^{ceil\left(\frac{N_{MO}}{s_2}\right)} \frac{s_2}{\sqrt{2\pi\sigma_2{}^2}} e^{\frac{(t-d_2 n)^2}{2s_2{}^2}} + \sum_{n=0}^{ceil\left(\frac{N_{FO}}{s_3}\right)} \frac{s_3}{\sqrt{2\pi\sigma_3{}^2}} e^{\frac{(t-d_2 n)^2}{2s_3{}^2}}$$

We decided to space homogeneous waves evenly, so each block of homogeneous waves can be modeled as a summation of normal distributions starting at regular intervals. We decided on four blocks of waves: one for premiers and pros, one for male open, and one for Athena/Clydesdale.

We separated the Athena and Clydesdale athletes from the other open registrants because their times were significantly slower, and put them both together as the last wave. The female and male professionals and premiers are together because they have very small populations. The first and last groups only have enough people for one wave, so for them no summation is necessary.

We plugged the standard deviations and mean times for the swim leg for these wave groups into our summation equations, and manually optimized the wave sizes and delays between waves to reduce maximum flow rate while not allowing the time required to release all waves to exceed the time we have available. We found that smaller waves released more frequently would lead to a lower max flow rate. We also found that because female opens have the smallest standard deviation, the females require more spacing between their waves than males. The wave sizes and delays that worked best were all 50 person waves with 41 seconds between male open waves, 40 seconds after the professional and premier wave, 53 seconds between female open waves, and 40 seconds before the Athena/Clydesdale wave.

Part II

Additional Assumptions

- **Rate does not change despite changing distances.**

After the creating a schedule for the race, we were asked to try and determine the benefits of altering the distances of the three race segments.

To accomplish this goal, we attempted to decrease the congestion at the first transition point as well as the amount of time contestants were running on the road. To do this we used the equation we found in part A:

$$p = \left(\frac{N_P}{\sqrt{2 s_1 \pi}} e^{-\left(\frac{(x-a_1)^2}{2 s_1}\right)}\right) + \left(\sum_{n=0}^{\text{ceil}\left(\frac{N_{MO}}{N_{H2}}\right)-1} \left(N_{MO} / (\sqrt{(2 s_2 \pi)})\right) e^{-\left(\frac{(x-nw_2-w_1-a_2)^2}{2 s_2^2}\right)}\right) +$$

$$\left(\sum_{n=0}^{\text{ceil}\left(\frac{N_{FO}}{N_{H3}}\right)-1} \left(N_{FO} / (\sqrt{(2 s_3 \pi)})\right) e^{-\left(\frac{\left(x-\text{ceil}\left(\frac{N_{MO}}{N_{H2}}\right) w_1-w_2-nw_2-b\right)^2}{2 s_3}\right)}\right) +$$

$$\left(\frac{N_C}{\sqrt{2 s_4 \pi}} e^{-\left(\frac{\left(t-w_1-\text{ceil}\left(\frac{N_{MO}}{N_{H2}}\right) w_2-\text{ceil}\left(\frac{N_{FO}}{N_{H3}}\right) w_3-w_4-a_4\right)^2}{2 s_4}\right)}\right) +$$

$$\left(\frac{N_A}{\sqrt{2 s_5 \pi}} e^{-\left(\frac{\left(t-w_1-\text{ceil}\left(\frac{N_{MO}}{N_{H2}}\right) w_2-\text{ceil}\left(\frac{N_{FO}}{N_{H3}}\right) w_3-w_4-a_5\right)^2}{2 s_5}\right)}\right)$$

But because we were altering the times, we had to alter the standard deviations.

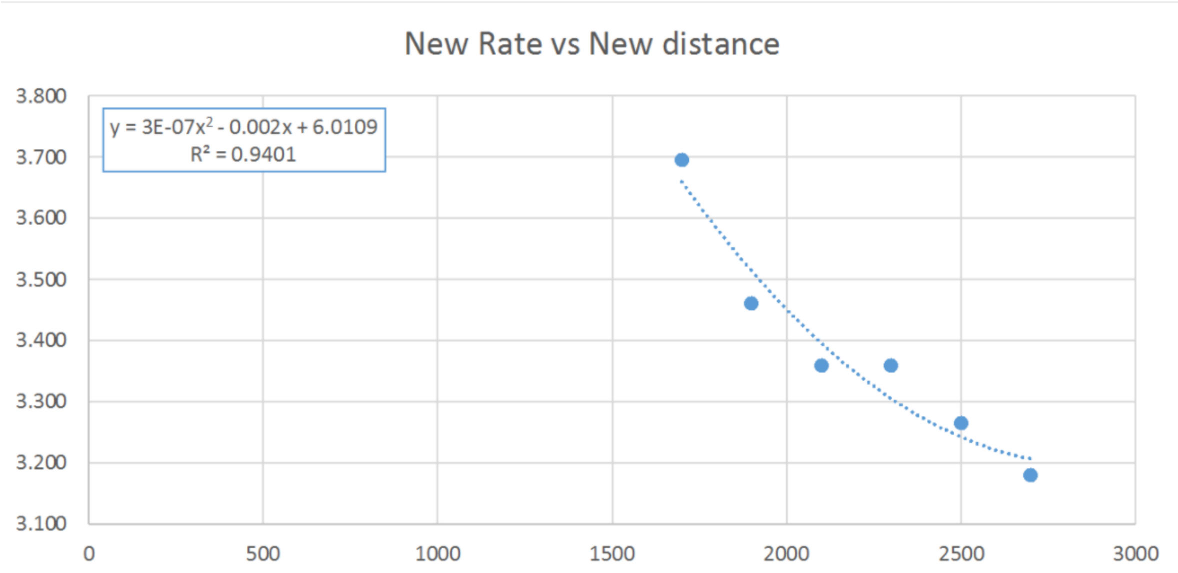
Every standard deviation was set equal to:
Sx = initial standard deviation of a given group
Dn = new distance
R = rate
Avgs = average swim time of a given group
Avg1 = transition time of a given group
New standard deviation = Sx * Dn * RAvgs + Avg1

The following table was composed to determine the rate of decrease of the function when the swim distance was increased.

In this table, new rate is the new amount of people coming out of the water every second if there are 50 people waves, and the time from the beginning of the race to when the last contestant leaves the start line is 12 minutes.

New Distance	New Rate	New Lt
1700	3.696	3:52:58
1900	3.462	3:52:02
2100	3.360	3:51:05
2300	3.360	3:50:09
2500	3.266	3:49:13
2700	3.180	3:48:17

The following graphs were made based on the table to determine a function which could determine rates without using the initial much longer equation.



The function is:

$y = (3*10^{-7})\, x^{2} - 0.002x + 6.0109$

After calculating the amount of congestion coming out of the water, we calculated the most amount of time that a single person would ever spend on land without changing the ratio of biking miles to was s while the ratio of biking miles to running miles was the same as the original ratio (40000:10000, 4:1). using the following equation:

Td = total distance (m)

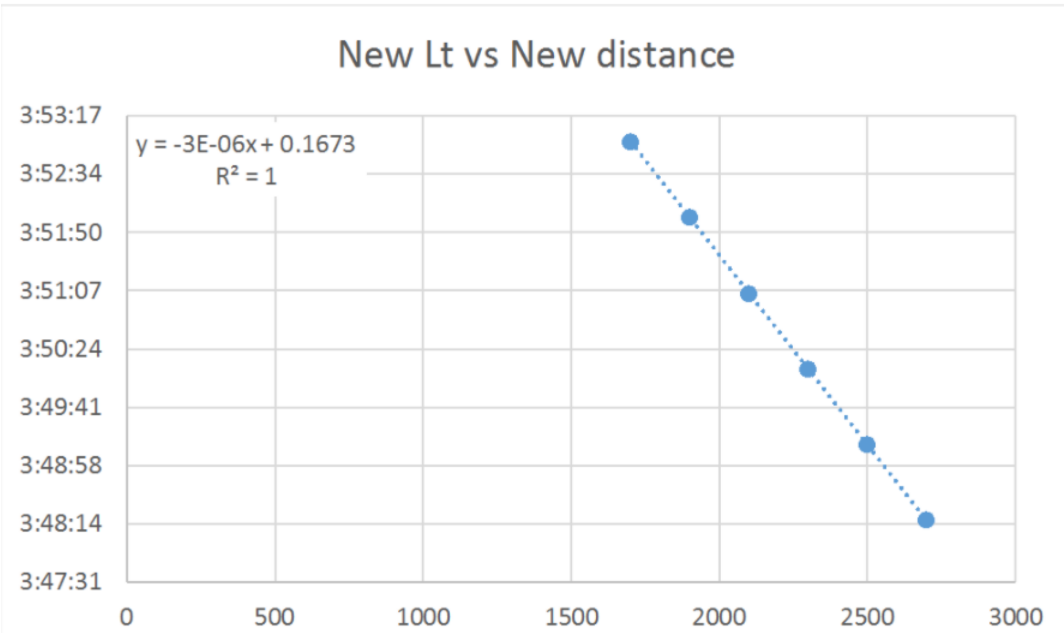
Sd = Swim distance (m)

Rb = Slowest biking rate (m/s)

Rr = Slowest running rate (m/s)

$$\frac{4(\frac{\mathrm{Td}-\mathrm{Sd}}{5})*\mathrm{Rb}+\frac{\mathrm{Td}-\mathrm{Sd}}{5}*\mathrm{Rr}}{86400} = \text{Total time (hours)}$$

The following graph is a graph of the longest amount of time a person would be out of the water compared to increasing the distance of the swimming portion.



This graph shows that the total amount of time decreases, therefore, if we were to increase the swimming distance the time we would need to close the roads would be decrease, as would the congestion at the beginning of the race.

Generalization = the more time spent in the water, the less congested the t1 will be. The body of water the athletes will be swimming in is wider than the streets where they will be biking and running. Therefore, it is most optimal to increase the time swimming because the athletes can spread out more in the water without having to worry about race congestion.

Analysis

Because we use only the mean, standard deviation, and skew of the data, our model takes the randomness of the data as a whole into account, as opposed to the min or max times which can vary greatly from event to event. The model accounts for the speeds of every athlete. We can easily plug different wave sets into our equation and see graphs of the outcomes.

Our model takes the distribution of athletes to be normal, and doesn't account for the fact that females, compared to males, have much faster swim times compared to final times. Therefore, many males will pass females on the biking leg, which our model doesn't consider. Our model assumes that we have enough beach area for 50 people to start side by side, and if there is less beach space then our model fails. It also doesn't optimize for any legs besides the swim.

Conclusions

To reduce the flow rate and congestion in a triathalon, we assumed that the swim times were the most important part and dictated the course of the rest of the triathlon. We assumed that the spread of athletes that formed during the initial section would carry over and cause minimal race congestion in other segments of the triathlon. However, individual result in each segment vary, and we didn’t account for individuals who performed really well in certain sections but not so well in others.

We can extend our equation to a computer program that would allow any athlete to be placed in any wave, and we can model the flow rates through all five checkpoints (swim, t1, bike..) and we could set up a computer to optimize all of them via brute force adjustment of the waves. We could also count the number of times throughout the race that any person passed anyone else. To do this we would count the number of pairs of people who end a leg in the opposite order than they started.