Optimize Road Closure Control in Triathlon Event

Zonghuan (Jason) Li
25 Longview Court, Hillsborough, CA 94010
jasonlzh2001@gmail.com

Mason Chen
2187 Morrill Avenue, San Jose, CA 95132
Mason.chen.training@gmail.com

Abstract
The Triathlon is an Olympic event that affects the entire city. This STEM paper is to manage the road closure time of a triathlon event in any City who may have a Triathlon event. In order to quantify the Road Closure Impact to meet 5.5 hours of closure time, authors have developed three different models: (1) Ideal Scenario, (2) Practical Scenario, and (3) Potential Scenario. In the first Ideal Scenario, authors applied Minitab descriptive statistics and box plots to optimize the event schedule and sequence among three Triathlon sports (Swim, Run, and Bike). In the second Practical Scenario, authors have managed the risks such as: pre-event preparation time, award ceremony time, dependency on Gender, Age, Category factors, 2000+ players, Safety/Health, etc. In the last Potential Scenario, we have further optimized the Zone Close and Open period. We can minimize the Zone wasted usage (waiting for players) based on the Worst Case Scenario. The biggest selling point of this paper is not just on the Closure Cost reduction, but also on the Safety/Health protection as well as promoting customer service and local business. This event may be completed on the event day, but our excellent customer service can attract more visitors to join the next city event.

Key Words: Triathlon, Descriptive Statistics, Minitab, Model, Non-Parametric

1. Introduction

Triathlon is a multisport event consisting of sequential swim, cycle, and run disciplines performed over a variety of distances[1]. The Purpose of this paper is to build an accurate model to manage a triathlon event in a city. There is one concern / restriction; the road closure time for the triathlon event has to be less than or equal to 5.5 hours in order to minimize the impact to the local people and business. We used data from 2017 HiMCM competition. In order to make that happen, we need to build a model that could tell us critical information like when we should hold the event, how are we going to arrange the sequence among the three sports etc. We will apply Minitab descriptive statistics, and three models to optimize the Road Closure time and area.

2. Building three Models

Three different models were constructed for three different scenarios; one being the ideal scenario, assuming that no unexpected events such as injuries, weather concerns occur. Therefore, the first ideal model may not be practical and needs certain improvements

The second model is the practical scenario, in which we consider the possible risks such as injuries, effect on local traffic and life overall, weather concerns, age dependency etc. We will take those into account and consider possible changes to the arrangement of the event in a practical sense.

The third model is the potential model, where we will figure out smarter ways to further optimize road closure time by sharing the streets between two events simultaneously to ensure we can finish the event and possibly add introduction and award ceremony and still being able to hold the road closure time to under 5.5 hours. In these models, Minitab software and descriptive statistics will be applied.

2.1 Proposal Strategy

We will try to minimize the road closure cost index (minimizing both the closure time and closure area), in order to ensure the minimum inconvenience to local residents and businesses. As priority, we will also take care of safety / health risks to ensure participants can feel safe. There is a broad range of possible medical problems and complications that must be taken into account when preparing for such races[4]. To add on to that, we will also provide world class costumer service to participants and the audience, by providing snacks, water etc. By doing this, we could also possibly help stimulate local business and find event sponsors, this gives us a opportunity to connect with them and possibly cooperate with them in the future.

2.2 Hypothesis

There are a few things we have to consider for hypotheses:

1, How exactly do we control the road closure time to be below 5.5 hours?
2, How do we accommodate more than 2000 Triathlon players in one event?
3, Can we take care of safety/health risks during the event?
4, How to we overcome unexpected weather situations?
5, What should the sequence between the three events be?
6, And do we start both genders together?

These are potential risks that, as a event planner, we need to take care of together.

3 Results

There are so many factors and constraints hidden in the project. There are so many alternative solutions which can resolve some problems if not all. There is no perfect scenario and all decisions are down to risk assessment and risk management. Based on the hypotheses, we run the raw data and build three models accordingly. The objective is to find a model which can meet 5.5 hours road closure time and also take care of safety and health concerns during the events.

3.1 Model 1: Ideal Scenario

This is an ideal model, assuming that everything runs perfectly and smoothly. First, we will apply one rule; we won’t invite any player who cannot finish the triathlon within 4:16:46 based on upper outlier criteria in order to control the total event time within 5.5 hours as shown in Figure 1. We used upper outlier criteria statistically in order to set a reasonable bar fairly.

Upper Outlier Criteria = Q3 -1.5* IQR IQR= Q3 – Q1
Q3= 75% Percentile
Q1= 25% Percentile

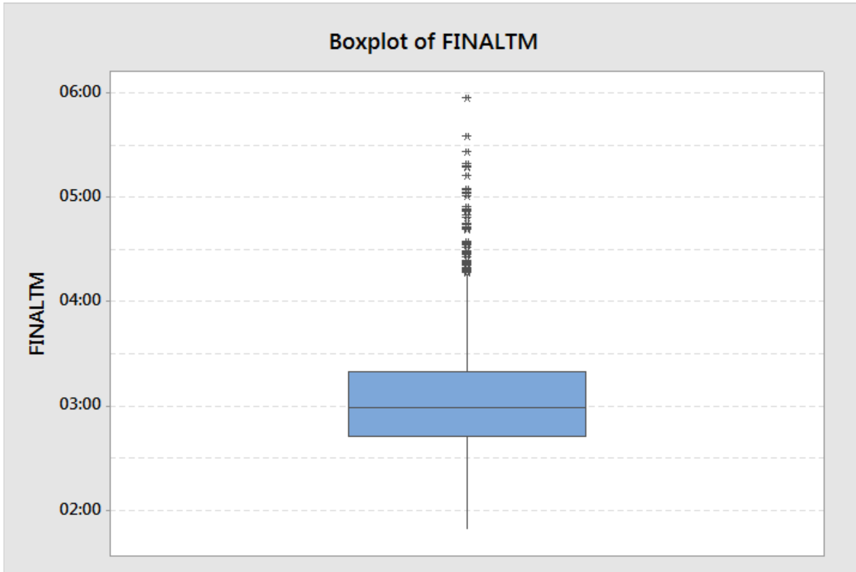

Figure 1: Boxplot for total time to finish the triathlon.

The reason why we chose a boxplot is because it shows us the distribution with many outliers. After applied Rule No.1, we re-plot the Boxplot as shown in Figure 2. We have excluded all the players with record beyond the upper outlier limit at 04:16:46. Total 59 players (2.8%) got impacted. The new distribution looks a normal distribution.

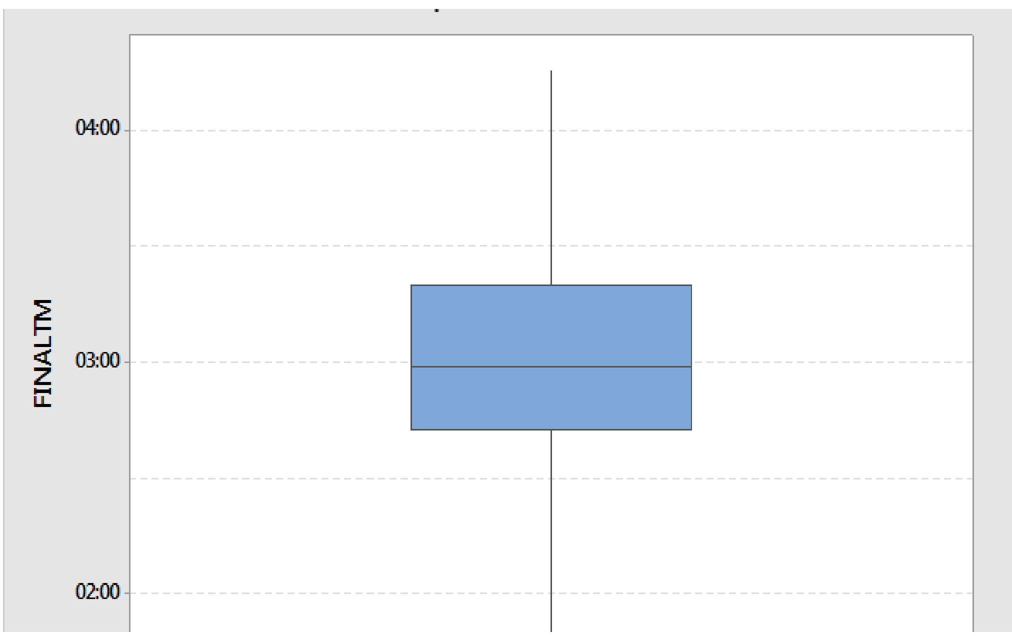

Figure 2: Boxplot for total time to finish triathlon after excluding outliers.

However, we cannot just ignore the 2.8% players that got impacted because they are slower players. In this model, we will leave that for now as a ideal scenario, and it will be discussed in model 2 practical model.

The next step is to determine how will we arrange the sport sequence. We need to make sure there is no significant overlap between the Bike and Run for safety reason if any collision risk happened between Bike Player and Run Player. We will first look at the distribution of three sports. As shown in Figure 3 Boxplot, we have observed the individual distribution of three sports. The Bike event has both the highest mean and the widest range.

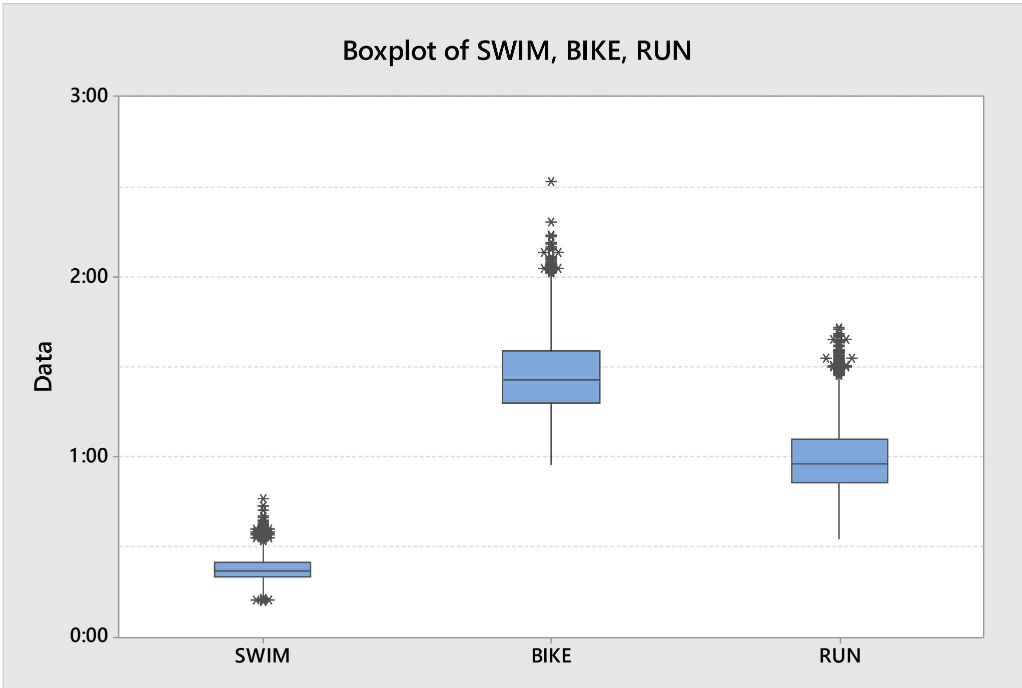

Figure 3: Boxplot on distribution of swim, bike and run.

This boxplot shows the time distribution of the three sports, as swim takes the shortest time and bike takes the longest. We should start Swim Event since the event is happened in the Ocean. We should take shorter Run event first over the longer and wider Bike event. Also, Bike event will take larger closure area. Therefore, the Bike event should be after the Run event.

We will also apply the second rule to disqualify any player who cannot complete both Swim and Run event within 01:48:32 based on the upper outlier criteria. Similar to first rule, the second criteria is necessary in order to ensure the entire road closure time is within 5.5 hours.

3.2 Model 2 Practical Scenario

This is the practical model that takes in more consideration on whether we should reserve any event preparation time in the beginning and hold an award ceremony in the end; how to handle players impacted by the rules listed in previous model 1? And also should we separate male and female participants to shorten the total road closure time?

We constructed the Non-Parametric Mann-Whitney Test for Median as shown in figure 4 since our raw data distribution is not normal and we have too many outliers which will distort the mean location. The reason why it is not normal is because pacing strategies during triathlon are highly influenced by distance and discipline[3].

	N	Median
FINALTM_F	743	0.12845
FINALTM_M	1894	0.12738

Point estimate for $\eta 1 - \eta 2$ is 0.00089
95.0 Percent CI for $\eta 1 - \eta 2$ is (-0.00036,0.00218)
W = 1004586.0
Test of $\eta 1 = \eta 2$ vs $\eta 1 \neq \eta 2$ is significant at **0.1625**
The test is significant at 0.1625 (adjusted for ties)

Figure 4: Results from Minitab software.

This shows that there is no significance (p value significance is above 0.05) on separating genders in the Triathlon event.

We also breakdown event time into 3 sports and the total time, Generally, time differences between sexes in swimming have been shown to be smaller on average than during biking and running[2]. Biking event has shown more significant gender dependence. We may consider separating the bike events between male players and female players. Though, which can create another management challenges during running-biking transition. Therefore, we won't recommend gender factor in our Triathlon event.

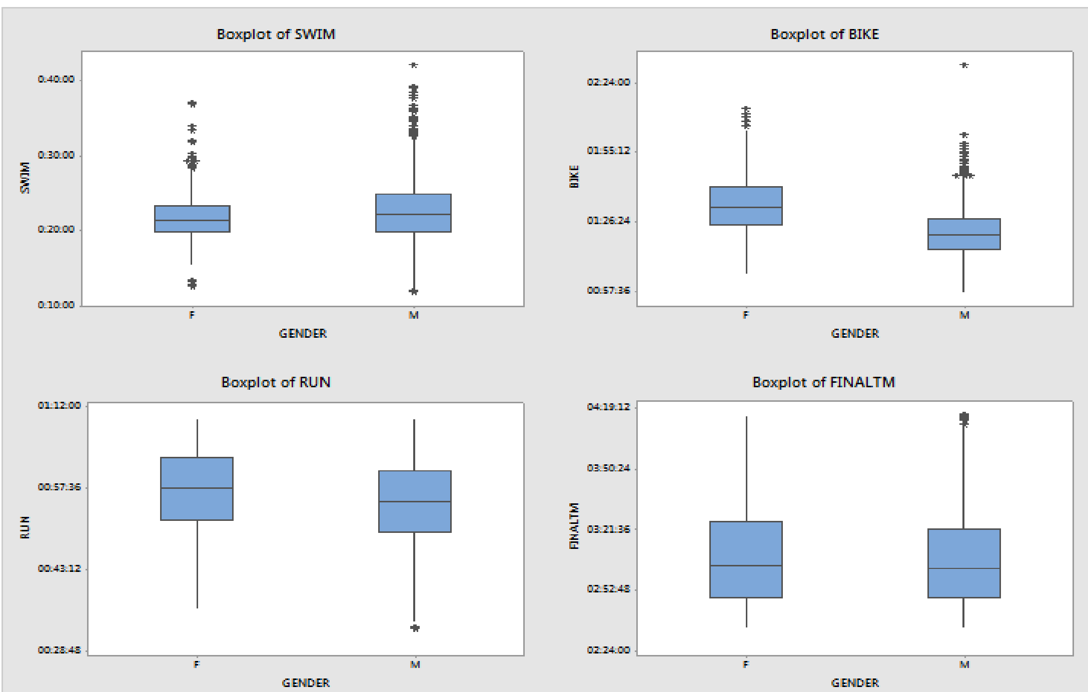

Figure 5: Boxplot on the time distribution for each sport and final time for both male and female participants.

These boxplots showed that there is no significant difference between female and male participants, so there is no point on separating them.

3.3 Model 3 Potential Scenario

Model 3 is the potential model that considers:

- How do we further optimize road closure time?
- Should we place these senior players in the back line when starting the Swim event?
- Can we claim any age dependency?

In order to answer these questions, we constructed a scatterplot and regression model by Minitab software as shown in Figure 6.

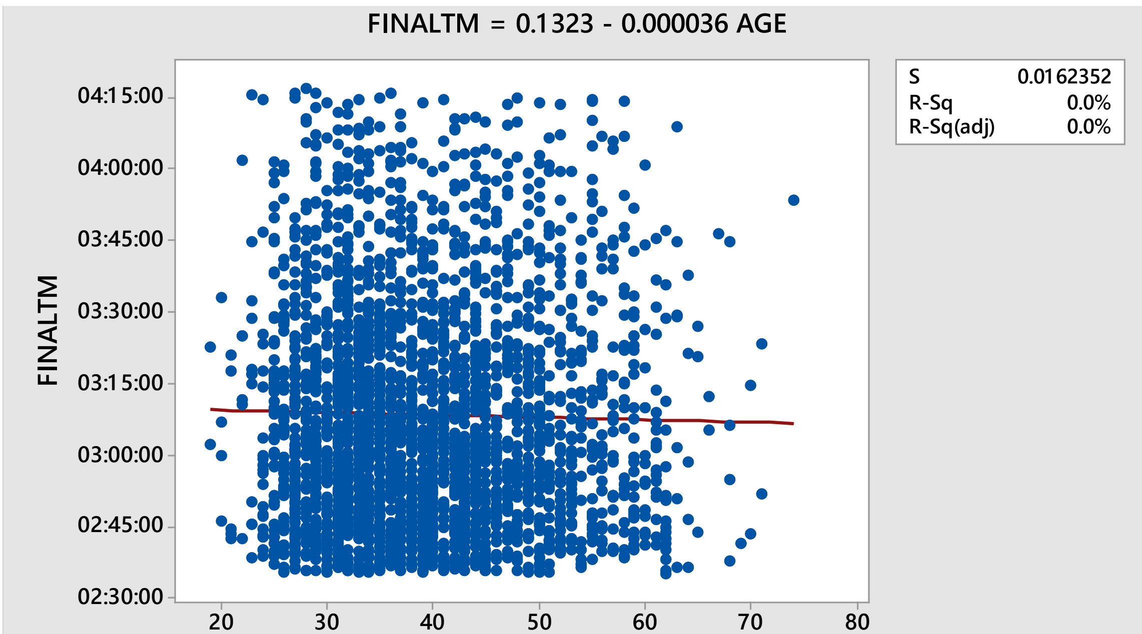

Figure 6: Scatterplot on age dependency.

We chose scatterplot because it may show us any correlation between the total time and age. As shown in the graph, the R-Square is only at 0% (random pattern), which means Age is not a factor that affects the participants' finish time. However, the level and age of triathlon competitors, and the race distance, influenced the risk of injury over a race series. These results provide timely information for triathlon race event organizers and could be incorporated into a review of practices for the provision of medical services to triathlon events, especially the common sprint distance competitions[5].

4. Conclusions

We will start Triathlon event with swim, then run, and bike based on the cycle time distribution and safety consideration. In order to meet 5.5 hours of road closure time requirement, we need to apply 2 disqualify rules to shorten the event duration. We will also hold the award ceremony in the end, and we will not separate male and female players or by age factor, as our model shows that there is no significance in doing so. We will take safety risk as top priority.

Acknowledgements

We will like to thank Dr. Charles Chen for Statistics Guidance. We will also like to thank our moms Linda Li and Chia Lin for supporting my STEM Projects.

Biography

Zonghuan (Jason) Li is currently a junior in San Mateo High School in San Mateo, CA. Jason moved to the U.S. from China in late 7th grade. During Jason's time here, he passed IBM SPSS statistics level 1, passed IBM SPSS modeler data mining v2, passed IBM SPSS data analyst v2. He passed AP Statistics, as well as got invited to several ASA conferences. Jason is familiar with Six Sigma DMAIC, Minitab, SPSS Statistics, SPSS Modeler CRISP Data Mining, AP Statistics, Programming, and ASQ (American Society for Quality) Quality Engineering. Jason also presented at the CSP conference in Jacksonville, FL and JSM Conference in Baltimore, MA. During his freshman year summer, he worked on a water purification lab with a professor and is currently working on publishing his paper; sophomore year summer, Jason served as a summer intern in one of Stanford University's bioengineering labs—Qi lab.

Mason Chen is a certified Six Sigma Black Belt, enrolled in the gifted Stanford OHS Program. Mason also certified IBM SPSS Statistics, Modeler Data Analyst and Data Miner. He has published 10 Conference Journal Proceedings Papers across his research in Computational Biology fields.

References

[1] Wu, Sam SX, Jeremiah J. Peiffer, Jeanick Brisswalter, Kazunori Nosaka, and Chris R. Abbiss. "Factors Influencing Pacing in Triathlon." Open Access Journal of Sports Medicine. Dove Medical Press, 16 Sept. 2014. Web. 23 July 2017.

[2] Lepers, R., B. Knechtle, and P. J. Stapley. "Trends in Triathlon Performance: Effects of Sex and Age." Sports Medicine (Auckland, N.Z.). U.S. National Library of Medicine, Sept. 2013. Web. 23 July 2017.

[3] Wu, S. S., J. J. Peiffer, J. Brisswalter, K. Nosaka, W. Y. Lau, and C. R. Abbiss. "Pacing Strategies during the Swim, Cycle and Run Disciplines of Sprint, Olympic and Half-Ironman Triathlons." European Journal of Applied Physiology. U.S. National Library of Medicine, May 2015. Web. 23 July 2017.

[4] Dallam, G. M., S. Jonas, and T. K. Miller. "Medical Considerations in Triathlon Competition: Recommendations for Triathlon Organisers, Competitors and Coaches." Sports Medicine (Auckland, N.Z.). U.S. National Library of Medicine, n.d. Web. 23 July 2017.

[5] Gosling, C. M., A. B. Forbes, J. McGivern, and B. J. Gabbe. "A Profile of Injuries in Athletes Seeking Treatment during a Triathlon Race Series." The American Journal of Sports Medicine. U.S. National Library of Medicine, May 2010. Web. 23 July 2017.