

For office use only

Team Control Number:

For office use only

T1 _____

#2016026 _____

F1 _____

T2 _____

F2 _____

T3 _____

F3 _____

T4 _____

F4 _____

IMMC 2016

The 2nd Annual International Mathematical Modeling Challenge Summary Sheet

(Attach a copy of this page to the front of your solution paper.)

Summary Sheet

Screaming at the top of our lungs, we eagerly witness world-class athletes set world record after world record from the spectators' stand. However, behind the adrenaline-induced races, gripping their armrests tightly and watching the races tensely, are the organizing committee and their insurance company -- the organizing committee bears the risk of paying a significant amount of money as a bonus to the winner if he or she succeeds in setting a new world record, while the insurance company may suffer a loss from insuring such a race.

In order to help the insurance company determine the price of the premium they offer that ensures their financial stability, and the organizing committee to make a the correct decision on when considering purchasing insurance, we looked into the complexity of the probability of record-breaking, premium calculation and decision making under risk in this paper, and came up with three different mathematical models targeting different scenarios when making the above decisions.

The most important factor affecting both parties' decisions is the risk of such an investment, which can be seen as the probability of an athlete breaking the world record. We developed a model using regression to analyze past data collected on athletes and their past records, then computed the standard deviation and normal distribution to predict the next record-breaking as a reference for both the insurance company and the organizing committee. The model was first built around one potential world record-breaker, and was advanced to accommodate multiple potential record-breakers. Since subsequent models are largely based on the result of this model, to access our model, we tested it to ensure its feasibility.

To help the insurance company determine the price of the premium, we developed another model that takes expected claim amount, safety-loading and profit-gaining loading, administrative costs, commission for agents, settlement of compensation and other factors into consideration. After deriving the equation to do so, we applied the eigenvector centrality concept to rank the importance of each factor.

Our third and final model helps the organizing committee decide whether to purchase insurance or not by considering the profit they gain after each competition and the likelihood of paying the bonus across a certain period. The model was later generalised so that it can be applied for multiple sport events using a single equation.

Afterwards we analyzed the strengths and weaknesses of our models, covering areas including but not limited to feasibility, efficiency and ease of usage, through which our team hopes to provide a clearer overview and further understanding of the models and each of their the merits and limitation. With real-life simulating models that consider a variety of scenarios and different factors, it is our sincere wish to provide models that are helpful and useable for the insurance company and the organizing committee to make better decisions for their own benefit.

Introduction

In athletics, athletes are often given a significant amount of bonus for breaking the world record of an event by the organizing committee of the competition, so as to attract top athletes to join said competition. As the risk of bearing a heavy financial burden if such an event takes place, the organizing committee has to decide whether to purchase insurance or not (in other words, self-insure). To aid the decision-making processes for both the organizing committee and the insurance company, we have developed several mathematical models that are presented in this paper, and test cases to assess the viability and sensibility of our models.

Problem Restatement

This year's problem requires us to do 5 things:

1. Calculate the average cost of the bonus for the Zevenheuvelenloop.
2. From the insurance company's perspective, determine the insurance premium for the Zevenheuvelenloop by considering different factors such as the average cost of the bonus, operating costs, time value of money, etc.
3. From the organizing committee's perspective, a) identify the criteria used to determine the necessity of purchasing the insurance for the Zevenheuvelenloop, and b) determine whether or not they should.
4. From the organizing committee's perspective, determine the way the organizing committee should weight each factor in deciding whether they should purchase the insurance for each of 40 events.
5. From the organizing committee's perspective, develop a mathematical model for the decision making process.

Overall Assumptions and Justifications

Assumption	Justifications
Accidents do not happen before or during the competition, i.e. the runners do not get injured or quit the competition under any circumstances once they are enrolled in the competition.	This is for the ease of calculation and to simplify the problem.
The performance of the runners are consistent and do not deviate too much from his or her average performance level.	
All runners engaged in the race aim to break the world record.	
Each competition's result is independent of others.	
Peak period of runners is similar for all athletes participating in the same event.	This assumption is made with references to past medical research ¹ on the peak periods of runners, muscle mass declinations and other factors.
Only the new record holder can receive the bonus.	This is for the ease of calculation and to simplify the problem, which also resembles real life situations where the organizing committee will do so to avoid a large financial burden.

¹ Schulz, R., & Curnow, C. (1988). Peak Performance and Age Among Superathletes: Track and Field, Swimming, Baseball, Tennis, and Golf. *Journal of Gerontology*, 43(5).

The first few records set in the earliest races are excluded from consideration during calculation.	Since the competitions might not have gained enough popularity to attract top runners, the records are not valid for projecting the probability of breaking records in the future.
All the money terms and prices in the paper are calculated in real terms without the effect of inflation.	This is to simplify our calculation and ensure the model can be used in the following years to come.
The year of usage of this model is seen as year 0 (or the 0 th year).	The model is designed to be used for not only this year, but anytime in the future; seeing the year of usage of the model as year 0 will allow calculations to be simplified.

Definition of Important Terms

Terms in the paper	Definition (explanation)
Dynamic information	Variables that change throughout the years
Constant	Variables that stay constant throughout the years
Top athletes (Potential winners)	Athletes who have the potential to break the world record
Peak age	Period in an athlete's life during which his or her performance is at the highest level
Average cost of the bonus	The ratio of the amount of bonus and the expected number of times the event is replicated before the current record is broken
Insurer	Insurance company, the party responsible for collecting the premium and determining the amount of the premium
Insured	The organizing committee in the competition in this paper, the party responsible for paying the premium and making the decision of whether to purchase the premium or not
Event	When the world record is broken by an athlete and the organizing committee has to pay him or her a bonus
Claim	The amount the insurer can claim back once accident occurs
Coverage period	The period in which the insurer is protected by the insurance plan

Models

1. Finding the expected average cost (Question 1)

1.1 Additional assumptions

Assumptions	Justifications
Top athletes have the chance of breaking the world record for more than once in successive years.	This is to simplify the problem and for the ease of our calculation
The performance of athletes obey the normal distribution.	
Chance of athlete breaking world record is exponentially proportional to the percentage difference of their personal best result and the world record.	
The average cost will not remain constant.	As society changes and technology develops, chances of nurturing top athletes also change and so does chance of breaking the world record.
The peak performance period of different athletes are the same.	This is to simplify our calculation and ensure the model can be used in the following years to come.

1.2 Addition Definition of Terms

Term	Definition
Average cost of year y	The cost times the chance of recording breaking in the said year

1.3 Method analysis

According to the definition of the average cost (the amount of bonus divided by the expected number of times the event is replicated before the current record is broken), since the former is fixed, the question then becomes finding the latter, which solely depends on the frequency/probability of breaking the world record. Therefore, we develop models to estimate the probability of the world record being broken.

Since the models of the following questions also rely on this probability, we have developed models that calculate not only the average cost of the 15K run as mentioned (i.e. Zevenheuvlenloop), but also that of all other athletics events. Two different models, the historical data model and the real time data model, are developed, and are optimised for different situations.

1.4 Historical data model

The historical data model directly calculates the probability of a top athlete breaking the world record using regression and analysis of historical data, and is best used for competitions that have few potential world-record breakers (for example, one in five years).

Data about past world records are collected, along with the year of achieving the record and the information of the athlete that set the record. To come up with a formula that can estimate when the next top athlete will be at peak age, By taking linear, exponential, logarithmic and quadratic regression on the data points of $h(x)$ = the year the x th world record was broken and using the one with the least error, we will be able to come up with a formula that can estimate when the next top athlete will be at the peak age.

Suppose the identified regression function is $f(x)$. Then the estimated time is $\mu = f(y + 1)$, where y is the number of data points currently on the graph. (Note: If more than one value needs to be estimated,

$f(y+2)$, $f(y+3)$ et cetera can also be calculated, but one should note that the error increases as the number of estimations increase.)

We will calculate the deviation of the regression (denoted by σ) of the derived function:

$$\sigma = \sqrt{\left(\sum_{i=1}^n |f(i) - h(i)|^2\right) \div n}$$

After the calculations of the desired values, we will progress to calculate the probability of the said top athlete breaking the record in each of the coming years. According our research², it is assumed that the performance of the athletes obey the normal distribution with mean at the peak age (that is, the top athlete has the highest probability of breaking the record at their peak age), the probability of the next top athlete breaking the world record in a certain year is:

$$q(x_0 : \text{date of calculation}) = \int_{x_0-0.5}^{x_0+0.5} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The reason why we can take the σ as the standard deviation in the above equation is that the physical meaning of σ is the error between the estimated peak age of past top athletes and their actual peak age; the more accurate our regression is, the more accurate our estimation of the peak age of athletes is (and thus a smaller σ), and the higher possibility the peak age of the future athlete can be estimated correctly, and vice versa. Thus, the error in our previous calculations can be used as the standard deviation in the above equation.

Then, the average cost of the competition for a year, x , will be $C \cdot q(x)$, where C denotes the prize money of the competition should the world record be broken.

1.5 Real time data model

The real time data model analyzes the probability of an athlete with a certain personal best record (PB) breaking the world record, and counts the number of athletes that stand a chance at breaking the world record and their respective probabilities, thus finding the overall probability of having at least one athlete breaking the world record. This model is optimised for events where there are frequently multiple top athletes competing every year.

We first narrow down the pool of athletes who apply to the event to so-called “top athletes”, those who stand a higher chance of breaking the world record, for the ease of calculation. This is done by comparing their current performances to the past performance of all world-record holders, as we believe them to be reasonable benchmarks for estimating an athlete’s future likelihood for breaking a record. The closer an athlete is to the benchmark, the higher his/her chance. Athletes whose performance is too far off stand a minimal chance and are filtered off, as the possibility of them improving to reach world-record standard in such a short period of time is close to zero. (Due to the nature of this model, this procedure has to be carried out only after the list of applicants has been confirmed, and so we can safely assume the period of time from the application and the actual event is too short to cause significant errors.) All that remains is to quantify the probabilities of the remaining potential winners breaking the record, then the expected value and thus average cost can be calculated.

² Schulz, R., & Curnow, C. (1988). Peak Performance and Age Among Superathletes: Track and Field, Swimming, Baseball, Tennis, and Golf. *Journal of Gerontology*, 43(5).