For office use only

T1____________
T2 ____________
T3 ____________
T4 ____________

Team Control Number

2016036

For office use only

F1 ____________
F2 ____________
F3 ____________
F4 ____________

2016
The international Mathematical Modeling Challenge (IM2C)
Summary Sheet

In athletic competitions, a big prize can be won if the world record is broken. In such events, the organizing committee will usually buy insurance since the financial risk can be quite large. Often, whether the insurance should be purchased or not may be difficult to determine.

To identify the decision on buying the insurance, we have developed two models. The Average Cost Model analyzes the chance of the record being broken at a particular edition. The Committee's Decision Model determines whether or not the organizing committee should purchase the insurance for each of the events.

The Average Cost Model uses the past records and world records of the event to analyze the chance of the record being broken. With the winning prize, the average cost per race can be calculated. It is a comprehensive approach to find the average cost, which can be used to determine approximately how much the committee is expected to pay for the winning prize per event. The chance of record being broken generated by this model is essential for the other models further on. The average cost of the 2016 edition of Zevenheuvelenloop is calculated as an example where all the actual records of the previous editions are used.

The Average Cost Model is advantageous due to its capability of predicting the probability of the record being broken at the current edition and the future editions from a small set of data by using simple calculations. However, curve fitting and integration are required.

The Committee's Decision model focuses on the offered insurance price and the probabilities of the record being broken in the future. The expected cost of self-insuring and being insured are compared to determine whether to buy the insurance or not. An offered premium price is generated as an example.

The Committee's Decision Model is advantageous due to its ability to calculate the threshold premium price. It can also be used with various events with different winning prizes and chance of winning. This model is also very simple when the data is analyzed from the Average Cost Model. It gives a definite result, which events to buy or not to buy. Nonetheless, it is assumed that the winning prize is the same for a number of editions. The accuracy depends on the chance of breaking the record. The lower the chance, the longer it will take for the record to be broken. Therefore, the more premium prices have to be calculated. A general decision-scheme is developed for the organizing committees to determine each event whether they should purchase insurance or self-insure.

From another perspective, the insurer also has to determine the price of the insurance. As a result, the Premium Price Model is developed. It focuses on the price they will be able to offer the insurance while covering all their expenses and also realize a profit.

Contents

1 Introduction

Setting the new world record time in a race is one of the most honorable awards that all athletes want to achieve. In the 15,000 meters race, the organizing committee offers a large amount of money in order to attract runners to their event. However paying the prize, 25,000 euro, could cause a financial problem for the organizing committee. Therefore, prize indemnity insurance might be required. Our models are designed to assess the chance of improving the record, determine the amount of expense to be added in average cost, deciding whether the organizing committee should purchase the insurance and choosing the essential events to purchase the insurance through identified factor in case there are numerous events.

2 Question 1

2.1 Problem Outline

The organizing committee is holding a running competition. To attract top runners, the committee would give 25,000 euro tothe runner who succeeds in breaking the world record. Although there is little chance for someone to break the record, the organizing committee willface a big financial problem if someone does.

The average cost is defined as the ratio of the prize by the expected number of repetitions before the record is broken.

Our task in this question is to develop a model that best predicts the chance of a runner breaking the record this year, which will be used to calculate the average cost. Prior data of this model includes skills and the amount of participants, which will be used in determining the average cost.

2.2 Assumptions

The following assumptions are taken into our consideration:

1. The prize for each repetitions is the same when the time passes
2. When generating histogram, the width of every intervals is set to 100s
3. The runners, those who were slower than the average, are not distributed in the graph since they have a lower chance of breaking the world record.
4. All runners are independent from one another.
5. These information are known:
    a. Previous data from the first repetition (including the finishing time of each runners)
    b. All previous world record

2.3 Average Cost Model

The Average Cost Model finds the average cost for a particular event. The chance of breaking the world record is calculated from past records.

2.3.1 Definition of Variables

P	Prize (Euro)
RB	The expected number of times the event is replicated before the current record is broken
WR	Set of previous world record time= $\{(x, y) \mid broken\ world\ record\ y(secs) in\ edition\ x\}$ e.g. let $x = 0$ represents 2000 Data: 1st world record 21 secs in 11th edition, 2nd world record 20 secs in 14th edition, 3rd world record 18 secs in 20th edition. $WR = \{(11,21), (12,20), (20,18)\}$
$t(i)$	Fitted Exponential function $[1/\exp(A \cdot i + B) + C]$ from WR
b_i	Chance of the record being broken in the i^{th} year
c_i	Percentage of the under graph area which is less than $t(i)$ in the i^{th} year
C	Average Cost (Euro)(P/RB)
p_i	Amount of participants in the i^{th} year
D_i	Set of finishing time of runners in the i^{th} year
y	Current edition e.g. 41st,57th

2.3.2 Algorithm

1. For $i = 1$ to $y - 1$
   a. Let F_i be the frequency distribution table from D_i with the width of the class interval of 100 seconds.
2. Let F_t be the summation of F_i for $i = 1$ to $y - 1$
3. Create histogram from F_t
4. For all $i \geq y$
   a. p_i will be the average from p_{i-1} to p_{i-5}
5. Ignore the data of runners those who ran not faster than the average.
6. Let $f(x)$ be the fitted Gaussian curve from F_tt
7. Let $t(x)$ be the fitted Inversed exponential function $[1/\exp(A \cdot x + B) + C]$ from WR, which represents the expected world record time that if to be broken in a year.
8. c_i isthe percentage of the under graph area which is less than $t(i)$ that represents the probability of people to break world record.

$$c_i = \frac{\int_{-\infty}^{t(i)} f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx}$$

It means that in every year c_i will change because $t(i)$ changes over time.

9. $b_i$is the probability of at least one person breaking the world record.

The probability of at least one person breaking world record is 1 – the probability of no one breaking world record which is $1 - c_i$ (when considering 1 person). Due to the runners are independent to each other, there are p_i participants, the probability of no one breaking world record is $(1 - c_i)^{p_i}$. Therefore, $b_i = 1 - (1 - c_i)^{p_i}$

10. So
    a. $RB = \frac{1}{b_y}$
    b. $C = \frac{P}{RB}$

2.4 Example

In this example, supposedly we are in the 33rd edition of theZeveheuvelenloopevent and we have the data from all those previous years. The winning prize is set to 25000 euro. We also know when the record is broken in the past.

So $P = 25000$, $y = 33$ and

$$WR = \{(12{,}2542), (14{,}2540), (15{,}2524), (18{,}2489), (27{,}2473)\}$$

Running recordsfrom years 2009(26th edition) to 2015(32nd edition)of the Zevenheuvelenloop were obtained from a website. Then we madea frequency distribution table with the width of interval class of 100s using Microsoft Excel.*[See Appendix Table A]*

The mean of the recordsobtained is 4732 seconds. There willbe 18,234 participants in the event. So $p_{33} = 18234$

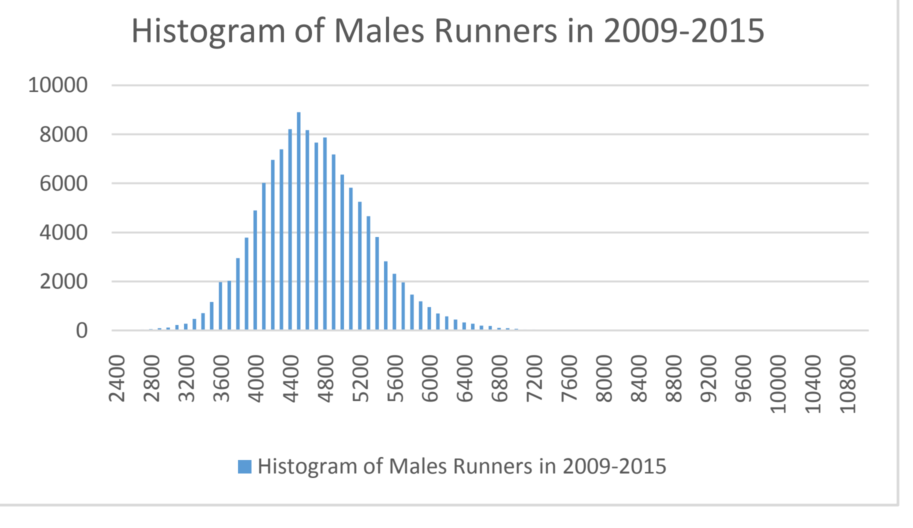

Figure 1 *Histogram of the 26th-32ndZevenheuvelenloop 15k males' running results*

From the assumptions (3), we have truncated the data point for those who ran slower than the average. Then, we fit the Gaussian distribution function using Logger Pro. This is the Gaussian distribution function.

$$g(x) = A \cdot e^{-\frac{(x-B)^2}{C^2}}$$

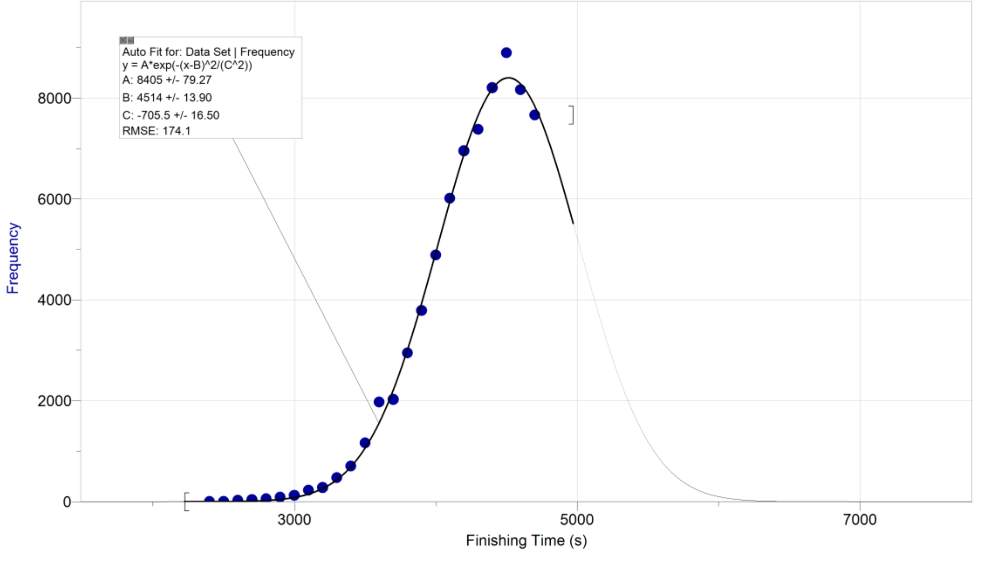

Figure 2 *The truncated histogram fitted with the Gaussian distribution function. (The blue dots represent the data from frequency distribution table (F_t). The black line represents the fitted Gaussian distribution function ($g(x)$))*

$$g(x) = 8405 \cdot e^{\frac{(x-4514)^2}{(-705.5)^2}}$$

Next, we will find $t(x)$ from $WR$with Inversed exponential function $[1/\exp(A \cdot x + B) + C]$using Logger Pro.

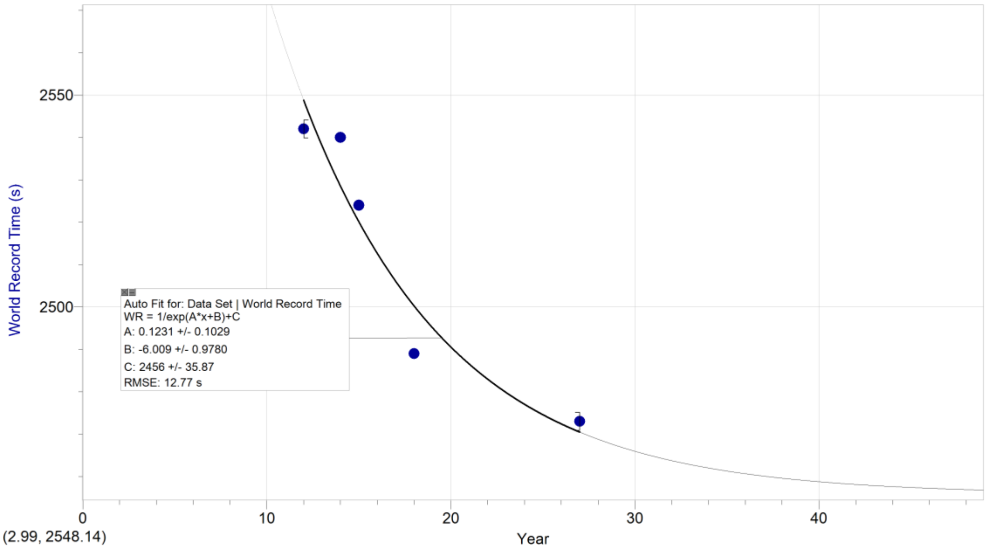

Figure 3 *The fitted graph of the all previous world records (The blue dots represent the record being broken, the black line represent the fitted curve.)*

$$t(x) = \frac{1}{e^{0.1231x-6.009}} + 2456$$

Then we find the probability of an average runner breaking the world record, which is calculating the ratio of the area beyond the world record time and the total area (c_i). The calculation can be done by integrating the $f(x)$.

The following is the integration of $f(x)$.

$$\int 8405e^{\frac{(x-4514)^2}{-705.5^2}}\,dx = -5.22508 \cdot 10^6 \operatorname{erf}(6.3983 - 0.00141743x) + constant$$

$$c_{33} = \frac{\int_{-\infty}^{t(33)} f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx} = 1.96692 \cdot 10^{-5}$$

Then we calculate b_{33}, which is the probability of at least one person breaking the world record.

$$\begin{aligned} b_{33} &= 1 - (1 - c_{33})^{p_{33}} \\ &= 1 - (1 - 1.96692 \cdot 10^{-5})^{18234} \\ &= 0.301 \end{aligned}$$

So

$$RB = \frac{1}{b_{33}} = 3.32$$

$$C = \frac{P}{RB} = 7530$$

Therefore, the average cost is 7,350 euro.

2.5 Strength and Limitations

The major strength of this Average Cost Model is that the Gaussian function fits the data very well. The histograms of every year have been observed. It is found that they are very similar to one another. This would represent the participants of each year accurately.So this model requires at least one year of past results and the all world records.

Also, the $t(x)$ function decreases if the record does not break, which means the chance of breaking the record is lower. Moreover, the $t(x)$ function can predict that probability for the future years.

One limitation of this model is that the$t(x)$ function has an asymptote, sothe world record time cannot be less than 2456 seconds, which is not quite realistic.

Another limitation is that this model cannot take into account the participants attending this year. For example, if Usain Bolt or top Marathon runner is attending the event, the model will not consider that as a higher risk.

Lastly, integrating the Gaussian function requires some tools and would outcome in unusual way. For example, the outcome of integrating Gaussian function contains an erf function which could take us one more step calculating that erf function.

3 Question 2

3.1 Problem outline

The insurance company offers the prize indemnity insurance to the organizing committee of the event in order to transfer therisk of loss from one entity to anotherequitably, in exchange for premiums. A premium consists of the claim (average cost) and additional costs.The insurance company will obtain the premium price that covers all profit and cost that the company needs to pay rationally.

Our goal is to find the reasonable amount of the addition to addwhich will have to cover the operating cost and realize profit. The addition is determined individually for different events with the concern on the winning prize and chance of breaking world record in that event.

3.2 Assumptions

1. The prize for each repetition is the same for every edition.
2. These information are known:
   a. The winning prize
   b. The number of contestants in the event

c. The time records of the runners in the past events
3. Time value is concerned
4. The scenario is considered under the present knowledge of training, material (shoes, suits and equipment), and drug laws.

3.3 Premium Price Model

This model is developed for insurance companiesto determine their premium price. It will cover all of the operating costs and the profit for the company. The result of the model will be the premium price offered to a customer (organizing committee).

3.3.1 Definition of variables

a_i	The percentage of the predicted premium price at year i^{th} when compared to the premium price at present time $a_i = \frac{b_i}{b_y}$
FOC	Fixed Operating Cost
n	Amount of subscribers of the insurance company
add	Additional cost that cover Fixed Operating Cost and profit such that Premium is $C \cdot (add + 1)$
P	Prize (Euro)
RB	The expected number of times the event is replicated before the current record is broken
C	Average Cost (Euro)= $\frac{P}{RB}$
T	Time Value Coefficient e.g. Economic Growth

3.3.2 Algorithm

Premium is $C \cdot (add + 1)$, that means premium is directly proportional to average cost. addin this term includes FOC, profits and probability to break world record. Factors that we are considering include the operational cost, the reasonableness of the premium, the profit, and the time value.

A complete algorithm is used here to determine the premium price:

1. Predict the RB of that event from the past statistics of that eventand number of contestant in that year of this event (Average Cost Model)
2. Calculate the average cost (C) of that event from RB
3. Determine the value of add with theconstraints (prioritized from high to low)
    a. $add \geq \frac{FOC}{n \cdot C}$
    b. To maximize the profit with the customer still wants to purchase, the following constrains is set. (If add also follows the constraint.)
        i. Using the Committee's Decision Model to predict $sumA$

$$C \cdot (add + 1) \leq \frac{P}{sumA}$$

$$add \leq \frac{P}{sumA \cdot C} - 1$$

c. If not, then it is unlikely to get an acceptance from the customer.
d. And the profit of the insurance company is

$$Profit = \frac{C \cdot add - \frac{FOC}{n}}{1 + T}$$

3.4 Example

In this example, we are the insurance company and are considering to offer an insurance for the 33rd Zevenheuvenlenloop running event. The winning prize is set to 25,000 euro. The company has 3,000 subscribers including this events with the operational cost of 300,000 euro per year.

So $P = 25{,}000, n = 3{,}000, FOC = 300{,}000$, and $y = 33$

By using the Average Cost Model we get the following for the current year ($y = 33$)

$\boldsymbol{i}$	$\boldsymbol{p_i}$	$\boldsymbol{t(i)}$	$\boldsymbol{c_i}$	$\boldsymbol{b_i}$	$\boldsymbol{RB}$
33	18234	2463.0	1.97E-5	0.301	3.32

Then we get the average cost $C = 25{,}000/3.32 = 7{,}530$ euro.

First, we consider the first constrain $add \geq \frac{FOC}{n \cdot C} \rightarrow add \geq 0.013$

By using the Committee's Decision Model, we get the following results.

$\boldsymbol{i}$	$\boldsymbol{p_i}$	$\boldsymbol{t(i)}$	$\boldsymbol{c_i}$	$\boldsymbol{b_i}$	$\boldsymbol{a_i}$
33	18234	2463.0	1.97E-5	0.301	1.00
34	18152	2462.2	1.95E-5	0.298	0.99
35	18062	2461.5	1.94E-5	0.296	0.98

The a_i coefficient predicts the decay of the premium for RB years ahead, so the premium of year i is expected to be $a_i \cdot X$.

The threshold price, the maximum price that the Committee can give to us, can be calculated by dividing the prize by $sumA$ which is $25000/2.97 = 8412.7$ euro.

From the second constrain $add \leq \frac{P}{sumA \cdot C} - 1 \rightarrow add \leq 0.117$ which also follow the first constrains. So to maximize the profit that the Committee's still wants to purchase, we let add to be the maximum value possible which is $add = 0.117$.

The premium that we going to offer the customer is $C \cdot (add + 1) = 8412.7$ euro. We will get the profit of $C \cdot add - \frac{FOC}{n} = 781$ euro.

3.5 Strength and Limitations

The major strength of this model is that the insurance company can make the highest amount of money. It would not result as a loss, which means in the worst case the company neither gets profits nor loss.

The limitation of this Premium Price Model is that we are assuming that the customers are using the same model of prediction. Hence sometimes this model would not yield a maximum profit depending on the model the customer using, in case that the model of customer return the higher threshold premium than this model's.

4 Question 3

4.1 Problem outline

The organizing committee is conducting a running event with a winning price for the winner if he or she succeeds in setting a new world record. To attract top runners, the amount of money can be quite large.The committee can decide to take a prize indemnity insurance to insure the prize they have to pay if a runner succeeds in breaking the record.

In this problem, we are required to construct a model to decide whether or not the organizing committee should purchase the insurance.

The winning price, the expected number of times the event is replicated before the record is broken, the current insurance price offered by the insurance companyand the predicted insurance prices will be used to determine the decision.

4.2 Assumptions

The model takes into consideration the following assumptions:

1. The expected number of times the event is replicated before the record is broken is estimated to be a whole number.
2. The prize for each repetitions is the same when the time passes
3. There will not be more than 1 record breaking in RB years.
4. These information are known:
    a. The number of participants
    b. The time records of the runners in the past events
5. The scenario is considered under the present knowledge of training, material (shoes, suits and equipment), and drug laws.

4.3 Committee's Decision Model

We have developed this model to help the committee to determine whether or not they should purchase the insurance. The result of the model will be a simple yes or no for a particular event.

4.3.1 Definition of Variables

a_i	The percentage of the predicted premium price at year i^{th} when compared to the premium price at present time
b_i	The probability of the record being broken at year i^{th}
c_i	Percentage of the under graph area which is less than WR in the i^{th} year
RB	The expected number of times the event is replicated before the current record is broken
X	The offered premium price from the insurance company of the current year
P	Prize
D_i	Set of finishing time of participants in the i^{th}edition
$sumA$	$\sum_{i=y}^{y+RB_{e,i}-1} a_i$
$f(x)$	Frequency distribution of the finishing time (x) of participants
$t(i)$	The world record time to be broken at editioni^{th}

4.3.2 Algorithm

The predicted price of self-insuring and the predicted price of being insured by the insurance company are compared over a period of time to determine whether the committee should buy the insurance or not.

From the predicted RB at the current year and the winning prize, the predicted average price of self-insuring is the winning prize over RB.

The predicted price of being insured by the company can be calculated from the current premium price and the predicted a_i, which will be derived from the probabilities of the record being broken at year i.

A complete algorithm is used here to determine the committee's decision:

1. The offered premium price X is based on the insurance company
2. Create a list of the probabilities of the record being broken at year i (b_i) calculated in the previous model from $i = y$ to$y + RB - 1$
3. Create a list of the percentage of the predicted premium price at year i when compared to the premium price at year $y(a_i)$ from $i = y$to $y + RB - 1$ from the list ofb_i when$a_i = \frac{b_i}{b_y}$

4. Find the summation of the a_i from $i = y$ to $y + RB - 1$ Let it be $sumA$
5. Multiply $sumA$ with the offered premium price by the insurance company X
6. Compare the calculated price to the winning prize
    a. If the calculated price is more than the winning prize
        i. The committee should buy the insurance offered by the company
    b. Else
        ii. The committee should not buy the insurance

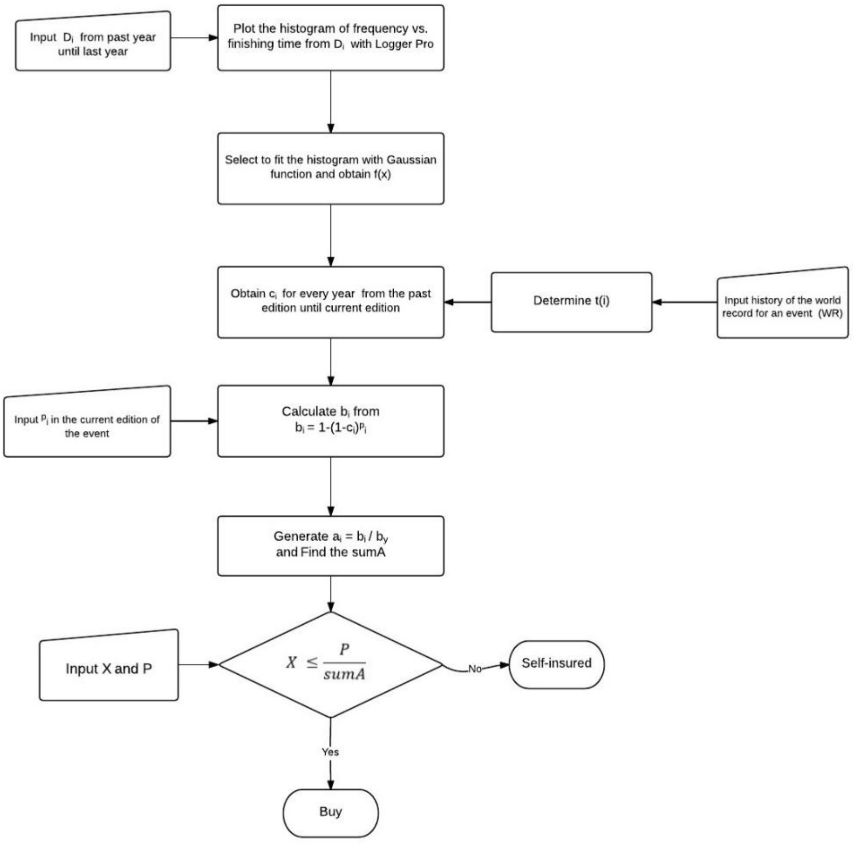

Figure 4*: the process of deciding what event should be insure for an event*

4.4 Example

In this example, supposedly we are conducting the 33rd edition of the Zeveheuvelenloop event. The winning prize is set to 25,000 euro and the insurance company offers a price of 8,000 euro.

So $P = 25000$, $X = 8000$, and $y = 33$

We will estimate the participant this and future year by calculating the average on 5 years before. For example, the number of participants of the 33rd edition comes from that of the 28th to 32nd edition.

i	p_i	i	p_i
28	18643	33	18234
29	18598	34	18152
30	18598	35	18062
31	17741	36	17955
32	17588	37	17998

By using the Average Cost Model we get the following for the current year ($y = 33$)

i	p_i	$t(i)$	c_i	b_i	RB
33	18234	2463.0	1.97E-5	0.301	3.32

From this point, we will predict RB years ahead (RB is rounded to the nearest whole number) to compare whether choosing an insurance service is better than self-insuring.

i	p_i	$t(i)$	c_i	b_i	a_i
33	18234	2463.0	1.97E-5	0.301	1.00
34	18152	2462.2	1.95E-5	0.298	0.99
35	18062	2461.5	1.94E-5	0.296	0.98

The a_i coefficient predicts the decay of the premium for RB years ahead, so the premium of year $i$is $a_i \cdot X$. By self-insuring, we expected to pay P/RB for every year.

i	**Self-insuring**	**Insured by the company**
33	8,333.33	8,000
34	8,333.33	7,923.5
35	8,333.33	7,850
Sum	25,000	23,773.5

We can clearly see that choosing the insurance company is the better choice as the expected amount to pay is less than by self-insuring.

To generalize, we calculate for $sumA$ which equals to $1.00 + 0.99 + 0.98 = 2.97$. The threshold price, the maximum price that we can give to the insurance company, can be calculated by dividing the prize by $sumA$ which is $25000/2.97 = 8412.7$ euro. That is we will accept the offer for any price less than or equal to 8412.7 euro. From this, we can see than the offered price X (8000) is less than our threshold. So the committee should buy the insurance offered by the company. If not, the committee should not buy the insurance offered by the company.

4.5 Strength and Limitations

The strength of this model is that it is relatively simple once we have done the process on the Average Cost Model.

Also the model is flexible as if the committee has the data for the past premium price, we could have used that data to predict the future premium price (a_i)

The limitation of this Committee's Decision model is that we are assuming that the record could be broken at most once in RB years. But in reality, there is a possibility that the record could broke twice in consecutive years.

5 Question 4

5.1 Problem outline

In case that the organizing committee sponsors many events, in order to reduce the risk of losing money, the organizing committee need to know which insurance to be purchase

Our model tries to find the most effective way of reducing risk. Hence the model will determine which insurances should be purchased to efficiently lower the risk.

Deciding which insurances should be purchased is determined based on many factors. These factors and also their effect will be indicated in term of weight.The significant factors are the past record of athletes in each event, the winning prize and the risk of the world record to be broken. The weights of each factor are determined in the algorithm of making the decision (as depicted in the model).

5.2 Assumptions

The model takes into consideration the following assumptions:

1. The expected number of times the event is replicated before the record is broken is estimated to be a whole number.
2. The prize for each repetitions is the same when the time passes
3. These information are known:
    a. The number of participants in each event that the organizing committee sponsors
    b. The list and number of event that the organizing committee sponsors
    c. The records of the athletes in the past events that the organizing committee sponsors
    d. All of the past world records

4. The scenario is considered under the present knowledge of training, material (shoes, suits and equipment), and drug laws.

5.3 Multiple Events

The previous model are used to find whether or not we should buy the insurance. For this question, we will just simply based on the Committee's Decision model on determining yes or no in every events.

5.3.1 Definition of Variables

$RB_{e,i}$	The expected number of times the event is replicated before the current record is broken of the event e in the i^{th}year
X_e	The offered premium price from the insurance company of the current year of the event i
WR_e	The set of past world record for event e
P_i	Prize of event i
$D_{e,i}$	Set of finishing time of participants in the i^{th}year of event e
E	Set of events that are insured by the company
a_i	The percentage of the predicted premium price at year i^{th} when compared to the premium price at present time
b_i	Chance of the record being broken in the i^{th}year
c_i	Percentage of the under graph area which is less than WR in the i^{th} year

5.3.2 Algorithm

1. For each event e
    a. Do the Committee's Decision Model with the following inputs:
        i. RB_i
        ii. X_i
        iii. P_i
        iv. $D_{i,j}$; $1 \leq j < y$
    b. if the result from the Committee's Decision Model is yes
        i. Add e to E
2. The committee should buy the insurance offered by the company for every event in E

5.4 Strength and Limitations

The strength of this procedures is that is you can determine whether or not you should buy an insurance for a particular event by looking at them individually.

The limitations of this model is that when there are some constrains, for example, the budget available, this model could not decide what events needs to be considered first and last.

6 Question 5

6.1 Problem outline

Deciding which insurance should be purchased for an event is determined based on the factor indicated in the previous question. The risk of the record to be broken (risk of losing money) and the value of prize in each event are mainly concerned. After consideration we concluded that both factors lead to the same objective, which is to reduce the amount of money that organizing committee has to pay. The same principle can also be applied for multiple eventsincluding long jump, high jump, etc. Thereforethe committee's decision model is used in this question again by considering each events separately and eventually making the decision.

6.2 Assumptions

The model takes into consideration the following assumptions:

1. The number of participants is equal in every year for a particular event in the future.
2. The prize for each repetitions is the same when the time passes
3. These information are known:
    a. The probability of the record being broken in that year of event thatthe organizing committee sponsors
    b. The list and number of event that the organizing committee sponsors
    c. The records of the athletes in the past events that the organizing committee sponsors
    d. All of the past world records
4. The scenario is considered under the present knowledge of training, material (shoes, suits and equipment), and drug laws.

6.3 Decision Scheme

This decision scheme is constructed for the Committee to decide whether or not we should purchase the insurance. This decision scheme provides you a comprehensive procedures that guide you to the results.

6.3.1 Definition of Variables

X_e	The offered premium price from the insurance company of the current year of the event e
WR_e	The set of past world record for event e
P_e	Prize of event e
$D_{e,i}$	Set of finishing time of participants in the i^{th}year of event e
E	Set of events that are insured by the company
$p_{e,i}$	Amount of participants in the i^{th} year of event e

$a_{e,i}$	The percentage of the predicted premium price at year i^{th} when compared to the premium price at present time of event e
$b_{e,i}$	Chance of the record being broken in the i^{th}year of event e
$c_{e,i}$	Percentage of the under graph area which is less than WR in the i^{th} year of event e
$sumA_e$	$\sum_{i=y}^{y+RB_{e,i}-1} a_{e,i}$
$f(x)$	Frequency distribution of the finishing time (x) of participants
$t(i)$	The world record time to be broken at editioni^{th}

6.4 Method and Application Used

1. Input the past record finishing time for the particular event e as sets $D_{e,i}$. The model need at least one edition of past statistic, the more of the edition the more accurate the model is.
    a. Create the histogram from the $D_{e,i}$(frequency vs. finishing time)
    b. Fit the histogram with the Gaussian function from the range $-\infty\ to\ mean\ of\ the\ finishing\ time$ and obtain $g(x)$

    By using Logger Pro 3.9

2. Input the past world record for event e (WR_e)
    a. Plot the point from theWR_e(world record time vs. year)
    b. Fit the point with the $[1/\exp(A \cdot i + B) + C]$and obtain $t(i)$

   By using Logger Pro 3.9
3. Calculate$c_{e,i}$ from

$$c_i = \frac{\int_{-\infty}^{t(i)} f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx}$$

   From $c_{e,y}\ to\ c_{e,y+RB_y-1}$ by using wolfram alpha software
4. Input the number of participant of the current year.
5. Generate $b_{e,i}\ from\ b_{e,i} = 1 - (1 - c_{e,i})^{p_{e,i}}$
6. Calculate the $a_{e,i}\ from\ \frac{b_{e,i}}{b_{e,y}}$ and find $sumA_e = \sum_{i=y}^{y+RB_{e,i}-1} a_{e,i}$
7. Input the offered premium price $(X_e)$from the insurance company of the current year of the event e and the Prize of event $e(P_e)$
8. Check the constraint $X_e \le \frac{P_e}{sum\ A_e}$
    a. If yes
        i. add e to E
    b. Else
        i. ignore e

Then repeat the same process from 1.to 8. with all e in E

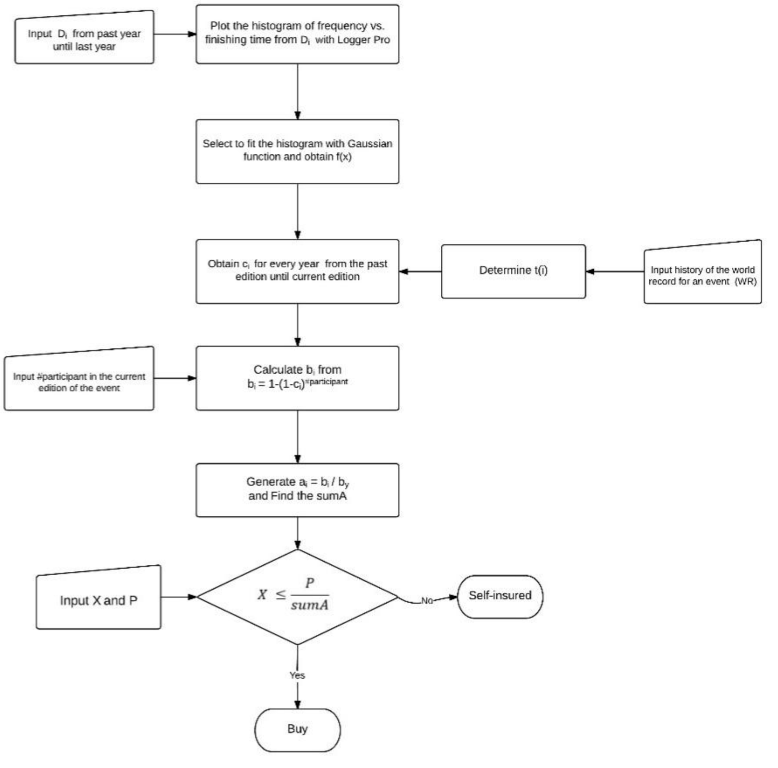

Figure 5*: the process of deciding what event should be insure for multiple events*

7 Further Development

7.1 Problem outline

Our previous attempt that tries to solve the question 4 and 5 is to determine what the insurances should be purchased if there are many events that organizing committee sponsors. Next the factors, which are important for the decision, are indicated and further weightedby its influence.

The concerned factor are probability of the record being broken (risk)and the cost of the winning prize. The main approach of this idea is trying to weight these factor individually not interpret both of them to be a value of money that the committee have to pay. This modelintroducednew idea called priority that can order the events in the order of insurance requirement.

7.2 Assumptions

The model takes into consideration the following assumptions:

1. The number of participants is equal in every year for a particular event from current edition and for the next $RB_{e,current\ \ edition}$ editions.
2. The expected number of times the event is replicated before the record is broken is estimated to be a whole number.
3. The prize for each repetitions is the same when the time passes
4. These information are known:
    a. The number of participants in each event
    b. The list and number of events that the organizing committee sponsors
    c. The records of the athletes in the past events (all of the events that organizing committee sponsors.) that will be used to predict the probability of the record being broken at present and also in the future
    d. The prize of each event
5. The scenario is considered under the present knowledge of training, material (shoes, suits and equipment), and drug laws.

7.3 Multiple Events Model with priority

The previous model are used to find whether or not we should buy the insurance. This model will just simply based on the previous model on determining yes or no in every events.

7.3.1 Definition of Variables

$RB_{e,i}$	The expected number of times the event is replicated before the current record is broken of the event e in the i^{th}edition
X_e	The offered premium price from the insurance company of the current year of the event e
P_e	Prize of event e
$D_{e,i}$	Set of finishing time of participants in the i^{th}year of event e
E	Set of events that are insured by the company
C_e	Average Cost (Euro)= $\frac{P_e}{RB_{e,y}}$ of event e
y	Current edition
$a_{e,i}$	The percentage of the predicted premium price at year i^{th} when compared to the premium price at present time
$sumA_e$	$\sum_{i=y}^{y+RB_{e,i}-1} a_{e,i}$
q_e	The priority of event e, which determine insurance requirement for that event by ${b_{e,i}}^2 \times P$
d_e	The deviation of the threshold prize($P_{threshold,e} = \frac{P_e}{sum\ A_e}$) of event e

7.3.2 Algorithm

The priority are identified as $chance^2 \times winning\ prize$ for two reason, first the probability of losing money influence our decision more than the winning prize and the second reason since $b_{e,i} \times P_e = C_e$ then the $q_e = C_e \times b_{e,i}$, which can be interpreted as the chance of losing average cost in an edition of event e

Next, the deviation of the threshold prize $d_e$is based on the idea that if the priority is higher then threshold offer price can be added the value in broader range. While if the priority is lower then threshold offer pricewill be added in the lower range. $d_e$can be simply expressed as $d_e \propto q_e\ or\ d_e = q_e\ \times constant(k)$

In thecommittee's decision model $P_{threshold,e} = \frac{P_e}{sumA}$but if the deviation is included then the $P_{threshold,e} = \frac{P_e}{sumA} + d_e$

The additional priority will be added in thecommittee's decision (multiple events) as shown below

1. For each event e
    a. Do the Model with the following inputs:
        i. $RB_{e,i}$
        ii. X_e
        iii. P_e
        iv. $D_{e,i}\ ; 1 \leq i < y$
    b. Calculateq_e
    c. Calculate d_e

    Repeat the same process in one for all of the $e\ in\ E$
2. Order the insurance in the arrangement of priority.
3. For each event e prioritized from high to low, compare and consider whether to purchase the insurance for that event or not with the modified threshold price.
    a. If $X_e\ \leq \frac{P_e}{sum\ A_e} + d_e$
        i. The committee should buy the insurance offered by the company
    b. else
        i. The committee should self-insured.
4. Continue considering the lower priority events until reach one of the constraints, for example limitation of the finance.

7.4 Strength and Limitations

This additional priority weight the chance more than the winning prize that have to be paid if the world record is broken. From the observation and researching the chance is found to be significantly more influence than the prize. However we cannot prove that this idea will effectively work in general cases. Another observation is that if the event has higher priority, that insurance tends to be purchased even if the prize is more than the original threshold price. Because the customers do not want to take a risk by themselves. In the original model, the committee's capital is assumed to be infinite. Nonetheless in reality the committee have to concern about their financial status so this priority will assist to identify what insurance are more influential. For the constant (k) are not yet indicated due to limitation of time, but one possible way that can help indicate this value is to collect the data and analyze through them. However, this model could give us inspiring idea that may have improved the current model.

8 Conclusion

In conclusion, we have proposed three models using previous records to calculate the average cost, the premium price and the decision for buying an insurance: first, the Average Cost Model, second, the Premium Price Model, and third, the Committee's Decision Model. We applied statistics and calculus to aid us in building the model to fit the graph and calculate the probabilities.

As of the questions, Question 1 is sufficiently addressed with the support of the Average Cost Model. From the records, the probability of breaking the world record is calculated. With the winning price of 25,000 euro, the average cost of the 15k run described in the problem can be calculated as 7530 euro. For Question 2, the Premium Price model is developed to find the premium price the insurance company should offer. For Questions 3, 4 and 5, the Committee's Decision Model can determine for each event whether they should purchase insurance or self-insure. There is also a general decision scheme, provided in Question 5, for organizing committees. Examples are displayed in each and every model for a more practical and clear understanding of the methods.

At last, the methods provided here can be applied to finding the probability of the record being broken in any event that involves breaking a record. There are still a number of assumptions in each of the questions that can still be generalized. In closing, the methods can be modified and developed further to introduce further restrains, improve its accuracy, capabilities and performance.

9 Appendix

9.1 Data Tables

Finishing Time	*Freq*	*Finishing Time*	*Freq*	*Finishing Time*	*Freq*	*Finishing Time*	*Freq*
2400	0	4600	8169	6800	108	9000	1
2500	1	4700	7665	6900	89	9100	0
2600	29	4800	7875	7000	69	9200	0
2700	41	4900	7177	7100	44	9300	0
2800	54	5000	6364	7200	40	9400	2
2900	86	5100	5819	7300	26	9500	0
3000	123	5200	5242	7400	16	9600	2
3100	225	5300	4662	7500	10	9700	0
3200	278	5400	3807	7600	16	9800	1
3300	473	5500	2816	7700	4	9900	0
3400	704	5600	2307	7800	8	10000	0
3500	1166	5700	1956	7900	3	10100	0
3600	1976	5800	1464	8000	2	10200	1
3700	2026	5900	1195	8100	0	10300	1
3800	2950	6000	956	8200	2	10400	0
3900	3791	6100	695	8300	2	10500	1
4000	4891	6200	572	8400	3	10600	0
4100	6016	6300	447	8500	1	10700	1
4200	6956	6400	327	8600	2	10800	0
4300	7383	6500	273	8700	3	10900	0
4400	8207	6600	192	8800	1	11000	1
4500	8901	6700	178	8900	0	11100	0

Table A Frequency Distribution table of the 26^{th} - 32^{nd}Zevenheuvelenloop 15k males' running results (F_t)

10 Bibliography and Tools

1. Wolfram Alpha[Computer software]. (2016). Retrieved from http://www.wolframalpha.com/
2. Logger Pro3.9 [Computer software]. (2016).
3. Zevenheuvelenloop website's Administrator. Retrieved May 9, 2016, from http://www.zevenheuvelenloop.nl/. Zevenheuvelenloop
4. Weisstein, Eric W. "Erf." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Erf.html
5. Weisstein, Eric W. "Gaussian Function." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/GaussianFunction.html