International Journal of Innovative Computing, Information and Control
Volume **14**, Number **5**, October **2018**
ICIC International ©2018 ISSN 1349-4198
pp. 1905–1915

MATHEMATICAL MODELS AT THE OLYMPIC GAMES TO PREDICT ROAD EVENTS

Marylú García Galván[1], Arnulfo Luévanos Rojas[1,*]
Sandra López Chavarría[1], Manuel Medina Elizondo[1]
José Benito Rivera Mendoza[1], Arturo Manuel Galindo Borrego[1]
Gabriel Aguilera Mancilla[2] and Jorge Luis Vázquez Lundez[2]

[1]Instituto de Investigaciones Multidisciplinaria
Facultad de Contaduría y Administración
Universidad Autónoma de Coahuila, Unidad Torreón
Blvd. Revolución 151 Ote. CP 27000, Torreón, Coahuila, México
*Corresponding author: arnulfol_2007@hotmail.com

[2]Facultad de Contaduría y Administración
Universidad Autónoma de Coahuila, Unidad Zona Norte
Carretera 57, km 4.5, CP 25710, Monclova, Coahuila, México

Received January 2018; revised May 2018

Abstract. *This paper presents five mathematical models at the Olympic Games to predict road events using regressions. The road events are: 1) Men's marathon; 2) Women's marathon; 3) 20 km Men's walk; 4) 20 km Women's walk; 5) 50 km Men's walk. The models have been developed through the data analysis of 18 Olympic Games from 1948 until 2016 for Men's marathon, 9 Olympic Games from 1984 until 2016 for Women's marathon, 16 Olympic Games from 1956 until 2016 for 20 km Men's walk, 5 Olympic Games from 2000 until 2016 for the Women's walk, and 10 Olympic Games from 1980 until 2016 for 50 km Men's walk according to the records. The road events are investigated by EXCEL to obtain the equations. The major contribution of this research is the proposal of the equations of each event to see if the athletes will break the records in the next Olympic Games. The equations that more are adjusted for the five road events are: the fifth-order polynomial equation fits the Men's marathon and Women's marathon; the second-order polynomial equation fits the 20 and 50 km Men's walk; the power equation fits the 20 km Women's walk. A comparison of the five models is presented between the proposed equations and real data. Also, a prediction is presented of the times for the Olympic Games of athletic in the road events for 2020 and 2024.*
Keywords: Mathematical models, Olympic Games, Predict road events, Fifth-order polynomial regression, Second-order polynomial regression, Power regression, EXCEL

1. **Introduction.** The modern Olympic Games or Olympics are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are held every four years, with the Summer and Winter Games alternating by occurring every four years but two years of apart [1].

The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilizing the mathematical models.

DOI: 10.24507/ijicic.14.05.1905

A number of researchers have studied a number of mathematical statistical models based on past performances in sports to predict future performances.

Prendergast applied the average speeds of world record times to determining a mathematical model for world records. The records or data used in the analysis spanned a 10 year period. Following his analysis raised the question of whether any further improvements can be expected or if the limits of human performance have been reached [2].

Heazlewood and Lackey proposed mathematical functions in the athletics for the men's events: for 100 m (inverse), 400 m (sigmoidal), long jump (cubic) and the high jump displayed four functions (compound, logistic, exponential and growth). In the women's events, the mathematical functions were 100 m (cubic), 400 m (sigmoidal), long jump (inverse) and high jump displayed four functions (compound, logistic, exponential and growth) [3].

Arnold and Godbey used basketball statistics to demonstrate the purpose of linear regression and to explain how to interpret its results. In particular, the student will quickly grasp the meaning of explanatory variables, r-squared, and the statistical significance of estimates of regression coefficients. Even if the student is not a sports fan the examples are easily understood and familiar. The student can easily replicate the procedures in this paper to reinforce learning [5].

Partington and Cushion investigated the coaching behaviors of elite English youth soccer coaches in different practice settings and gained insight into the coaches' cognitive processes underpinning these behaviors. The practice setting was split into two types of activities, "training form" and "playing form", and behavioral data were collected using a modified version of the Coach Analysis and Intervention System. Interpretive interview data were triangulated with the behavioral data to ensure that both the "what" and the "why" of the coaches' behavior and practice were considered [6].

Heazlewood made a comparison between the prediction data versus reality data, using mathematical models to predict elite performance in swimming and athletics at the Olympic Games [4].

Ahamad et al. identified essential parameters in sports for talent assessment in cricket. A model based on Ordered Weighted Averaging Aggregation (OWA) operator was proposed. This paper also presented an example demonstrating the application of the proposed algorithm on sample data [7].

Jackson introduced a measure of similarity between players in the Australian Football League (AFL). For each player, their signature is defined using a 175-dimension vector that represents a breakdown of a player's involvements during games. Each dimension is a unique combination of event type (43 classifications), event zone (three classifications) and game state (three classifications). The magnitude of each dimension is defined as the percentage of a player's total involvements that were observed for that particular combination of event, zone, and state [8].

Huynh and Marshall examined general tau theory in relation to the sport badminton. They cover the three main concepts of motion-gaps, tau-coupling, and time-to-contact. The second half of this paper examines how tau theory can be utilized by coaches and trainers in training their athletes. Specifically, they discuss anticipatory training, time-to-contact theory, skills acquisition training for badminton program, and the required velocity model [9].

Devoir et al. reported their recent findings on the analysis of riders' performance patterns in the new (third version) Track Cycling Omnium. They made use of statistical and machine learning techniques to analyze data in recent omnium competitions and compared the results with their previous findings on the older version of the omnium to understand how and whether the new omnium requires different skill sets and strategic planning for elite riders and their coaches [10].

O'Bree and Grundy developed a moving average based rating system using all tournaments from the 2012 through 2015 PGA Tour seasons. The system used a relative measure of performance taken to be the ratio of player round score to average field round score [11].

This paper presents five mathematical models at the Olympic Games to predict road events using regressions through the real data and by EXCEL are obtained the equations for Men's marathon, Women's marathon, 20 km Men's walk, 20 km Women's walk, and 50 km Men's walk. The above is to observe if the athletes of these road events will break the records in the next Olympic Games. Also, a comparison is presented between the proposed equations and real data to observe differences, and a prediction for the Olympic Games of athletic in the road events for 2020 and 2024 is shown.

The paper is organized as follows. Section 2 describes the formulas of the five mathematical models for predicting the times at the Olympic Games of athletic for the road events from historical data by EXCEL. Subsection 2.1 shows the model for the Men's marathon. Subsection 2.2 presents the model for the Women's marathon. Subsection 2.3 shows the model for the 20 km Men's walk. Subsection 2.4 shows the model for the 20 km Women's walk. Subsection 2.5 shows the model for the 50 km Men's walk. Results and discussion are presented in Section 3. Conclusions (Section 4) complete the paper.

2. Methodology.

The results for the finalists at the Olympic Games for the road events Men's marathon, Women's marathon, 20 km Men's walk, 20 km Women's walk, and 50 km Men's walk were collected from Internet-based results [12].

2.1. Modeling for the Men's marathon.

Table 1 shows the historical data of the times at the Olympic Games for the road events (Men's marathon) from 1948 until 2016 [12].

Games are considered from 1948 because games are canceled in 1940 and 1944 by world wars.

The road events (Men's marathon) of the last 18 Olympic Games from 1948 until 2016 from historical data by EXCEL the fifth order polynomial trend is obtained (see Figure 1). Figure 1 shows the historical data and the fifth order polynomial trend (equation).

In order to use the equation given in Figure 1, the following is done: the interval of each unit of the "x" axis is 1 and starts with 1. Then each interval is equal to four, because this event is realized each four years and unit 1 must be 1948, because the data is initiated from 1948, and thus ratio between the two variables should be $x = \left(\frac{x_1-1944}{4}\right)$.

Now, substituting $x = \left(\frac{x_1-1944}{4}\right)$ into equation shown in Figure 1 and the general equation of the road events (Men's marathon) is obtained:

$$y = 4(10)^{-5}\left(\frac{x_1-1944}{4}\right)^5 - 0.0004\left(\frac{x_1-1944}{4}\right)^4 - 0.0421\left(\frac{x_1-1944}{4}\right)^3 + 1.0986\left(\frac{x_1-1944}{4}\right)^2 - 9.8572\left(\frac{x_1-1944}{4}\right) + 162.27 \quad (1)$$

where y is the time in minutes of the road events, and x_1 is the year of the Olympic Games.

2.2. Modeling for the Women's marathon.

Table 2 presents the historical data of the times at the Olympic Games for the road events (Women's marathon) from 1984 until 2016 [12].

TABLE 1. Men's marathon

Olympic Games Year	Time Hours:Minutes:Seconds	Time Minutes
1948	2:34:51.6	154.86
1952	2:23:03.2	143.05
1956	2:25:00	145.00
1960	2:15:16.2	135.27
1964	2:12:11.2	132.19
1968	2:20:26	140.43
1972	2:12:19	132.32
1976	2:09:55	129.92
1980	2:11:03	131.05
1984	2:09:21	129.35
1988	2:10:32	130.53
1992	2:13:23	133.38
1996	2:12:36	132.60
2000	2:10:11	130.18
2004	2:10:55	130.92
2008	2:06:32	126.53
2012	2:08:01	128.02
2016	2:08:44	128.73

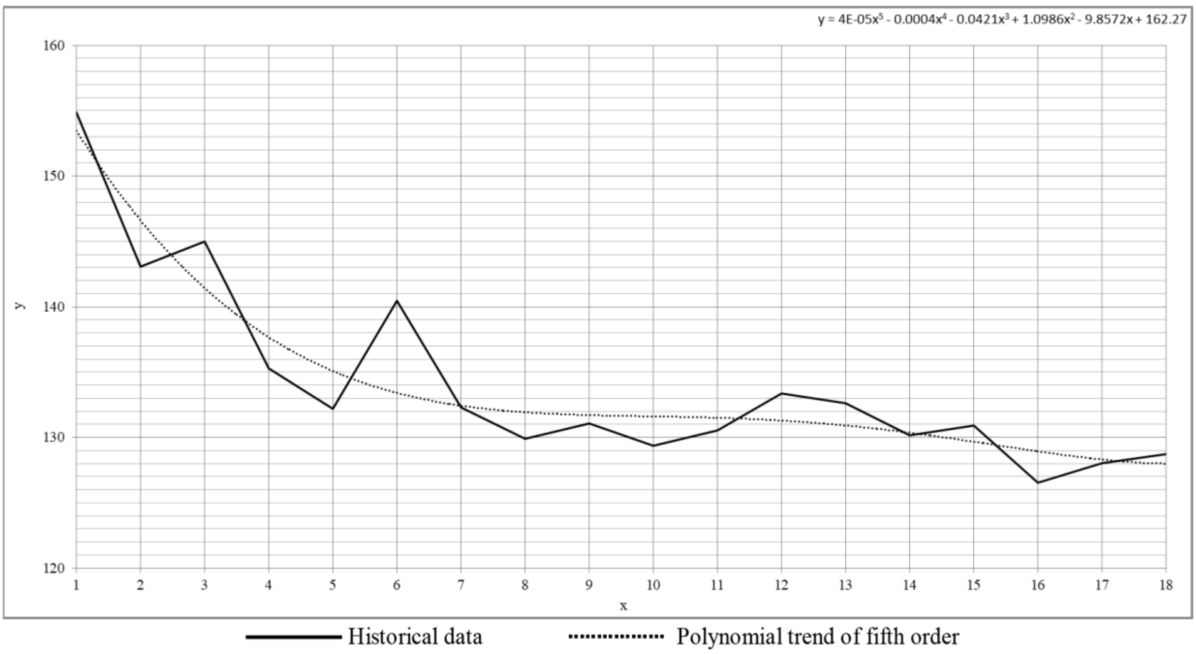

FIGURE 1. Men's marathon

Women's marathon is considered from 1984 because Women's marathon is incited in 1984.

The road events (Women's marathon) of the last 9 Olympic Games from 1984 until 2016 from historical data by EXCEL the polynomial trend of fifth order is obtained (see Figure 2). Figure 2 shows the historical data and the polynomial trend of fifth order (equation).

TABLE 2. Women's marathon

Olympic Games Year	Time Hours:Minutes:Seconds	Time Minutes
1984	2:24:52	144.87
1988	2:25:40	145.67
1992	2:32:41	152.68
1996	2:26:05	146.08
2000	2:23:14	143.23
2004	2:26:20	146.33
2008	2:26:44	146.73
2012	2:23:07	143.12
2016	2:24:04	144.07

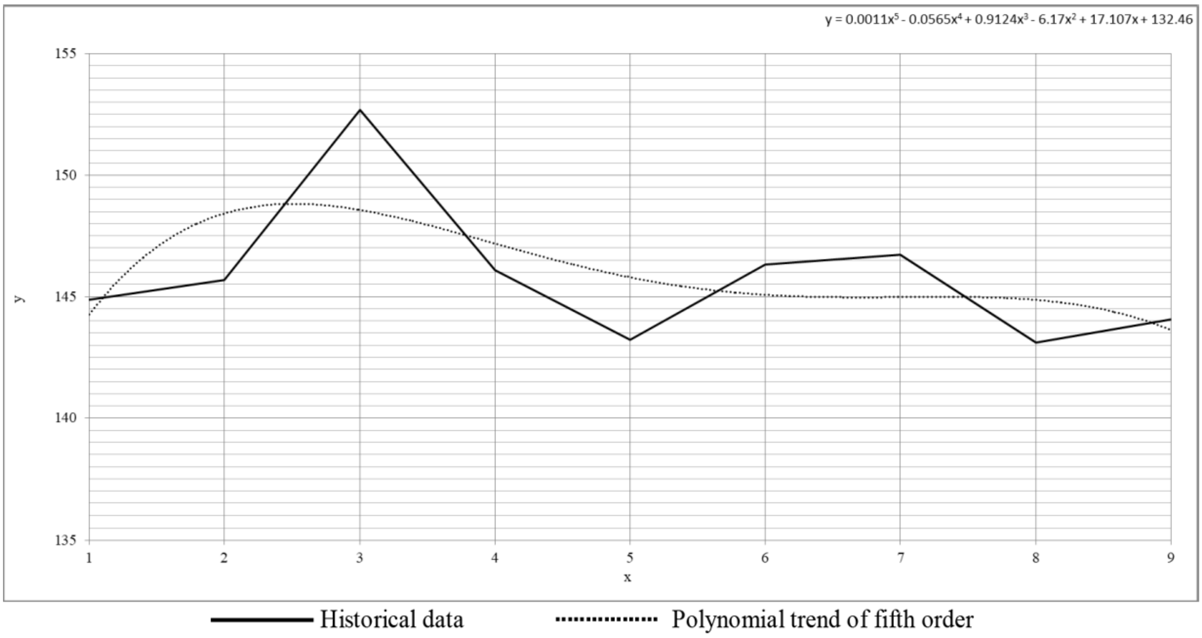

FIGURE 2. Women's marathon

In order to use the equation given in Figure 2, the following is done: the interval of each unit of the "x" axis is 1 and starts with 1. Then each interval is equal to four, because this event is realized each four years and unit 1 must be 1984, because the data is initiated from 1984, and thus ratio between the two variables should be $x = \left(\frac{x_1-1980}{4}\right)$.

Now, substituting $x = \left(\frac{x_1-1980}{4}\right)$ into equation shown in Figure 2 and the general equation of the road events (Women's marathon) is obtained:

$$y = 0.0011\left(\frac{x_1-1980}{4}\right)^5 - 0.0565\left(\frac{x_1-1980}{4}\right)^4 + 0.9124\left(\frac{x_1-1980}{4}\right)^3 - 6.17\left(\frac{x_1-1980}{4}\right)^2 + 17.107\left(\frac{x_1-1980}{4}\right) + 132.46 \quad (2)$$

2.3. **Modeling for the 20 km Men's walk.** Table 3 shows the historical data of the times at the Olympic Games for the road events (20 km Men's walk) from 1956 until 2016 [12].

20 km Men's walk is considered from 1956 because 20 km Men's walk is incited in 1956.

The road events (20 km Men's walk) of the last 16 Olympic Games from 1956 until 2016 from historical data by EXCEL the polynomial trend of second order is obtained (see Figure 3). Figure 3 shows the historical data and the polynomial trend of second order (equation).

In order to use the equation given in Figure 3, the following is done: the interval of each unit of the "x" axis is 1 and starts with 1. Then each interval is equal to four, because this event is realized each four years and unit 1 must be 1956, because the data is initiated from 1956, and thus ratio between the two variables should be $x = \left(\frac{x_1 - 1952}{4}\right)$.

TABLE 3. 20 km Men's walk

Olympic Games Year	Time Hours:Minutes:Seconds	Time Minutes
1956	1:31:27.4	91.46
1960	1:34:07.2	94.12
1964	1:29:34.0	89.57
1968	1:33:58.4	93.97
1972	1:26:42.4	86.71
1976	1:24:40.6	84.68
1980	1:23:35.5	83.59
1984	1:23:13	83.22
1988	1:19:57	79.95
1992	1:21:45	81.75
1996	1:20:07	80.12
2000	1:18:59	78.98
2004	1:19:40	79.67
2008	1:19:01	79.02
2012	1:18:46	78.77
2016	1:19:14	79.23

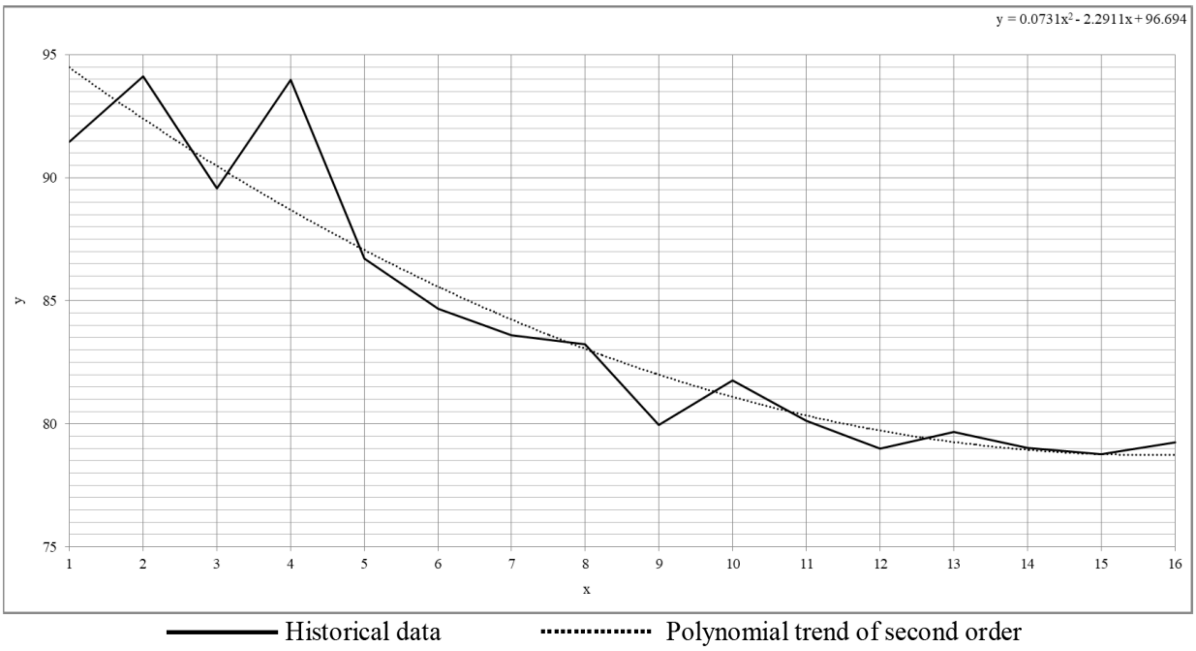

FIGURE 3. 20 km Men's walk

Now, substituting $x = \left(\frac{x_1-1952}{4}\right)$ into equation shown in Figure 3 and the general equation of the road events (20 km Men's walk) is obtained:

$$y = 0.0731\left(\frac{x_1 - 1952}{4}\right)^2 - 2.2911\left(\frac{x_1 - 1952}{4}\right) + 96.694 \qquad (3)$$

2.4. **Modeling for the 20 km Women's walk.** Table 4 presents the historical data of the times at the Olympic Games for the road events (20 km Women's walk) from 2000 until 2016 [12].

20 km Women's walk is considered from 2000 because 20 km Women's walk is incited in 2000.

The road events (20 km Women's walk) of the last 5 Olympic Games from 2000 until 2016 from historical data by EXCEL the power trend is obtained (see Figure 4). Figure 4 shows the historical data and the power trend (equation).

In order to use the equation given in Figure 4, the following is done: the interval of each unit of the "x" axis is 1 and starts with 1. Then each interval is equal to four, because this event is realized each four years and unit 1 must be 2000, because the data is initiated from 2000, and thus ratio between the two variables should be $x = \left(\frac{x_1-1996}{4}\right)$.

TABLE 4. 20 km Women's walk

Olympic Games Year	Time Hours:Minutes:Seconds	Time Minutes
2000	1:29:05	89.08
2004	1:29:12	89.20
2008	1:26:31	86.52
2012	1:25:02	85.03
2016	1:28:35	88.58

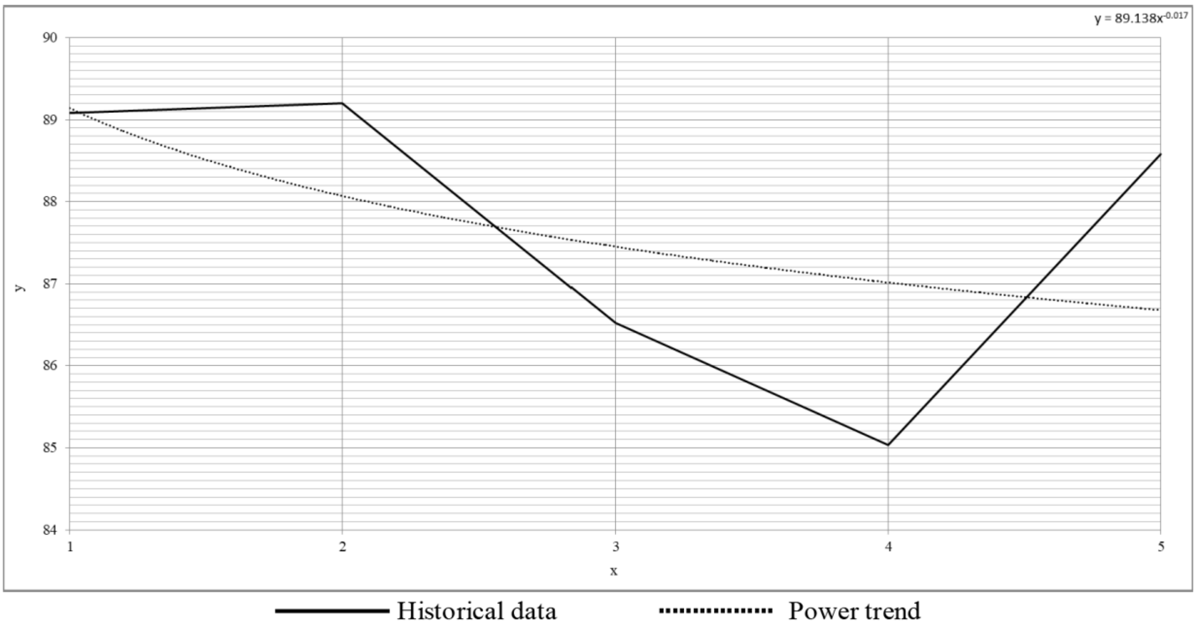

FIGURE 4. 20 km Women's walk

Now, substituting $x = \left(\frac{x_1-1996}{4}\right)$ into equation shown in Figure 4 and the general equation of the road events (20 km Women's walk) is obtained:

$$y = 89.138\left(\frac{x_1 - 1996}{4}\right)^{-0.017} \tag{4}$$

2.5. **Modeling for the 50 km Men's walk.** Table 5 shows the historical data of the times at the Olympic Games for the road events (50 km Men's walk) from 1980 until 2016 [12].

50 km Men's walk is considered from 1980 because 50 km Men's walk is canceled in 1976.

The road events (50 km Men's walk) of the last 10 Olympic Games from 1980 until 2016 from historical data by EXCEL the polynomial trend of second order is obtained (see Figure 5). Figure 5 shows the historical data and the polynomial trend of second order (equation).

TABLE 5. 50 km Men's walk

Olympic Games Year	Time Hours:Minutes:Seconds	Time Minutes
1980	3:49:24	229.40
1984	3:47:26	227.43
1988	3:38:29	218.48
1992	3:50:13	230.22
1996	3:43:30	223.50
2000	3:42:22	222.37
2004	3:38:46	218.77
2008	3:37:09	217.15
2012	3:36:53	216.88
2016	3:40:58	220.97

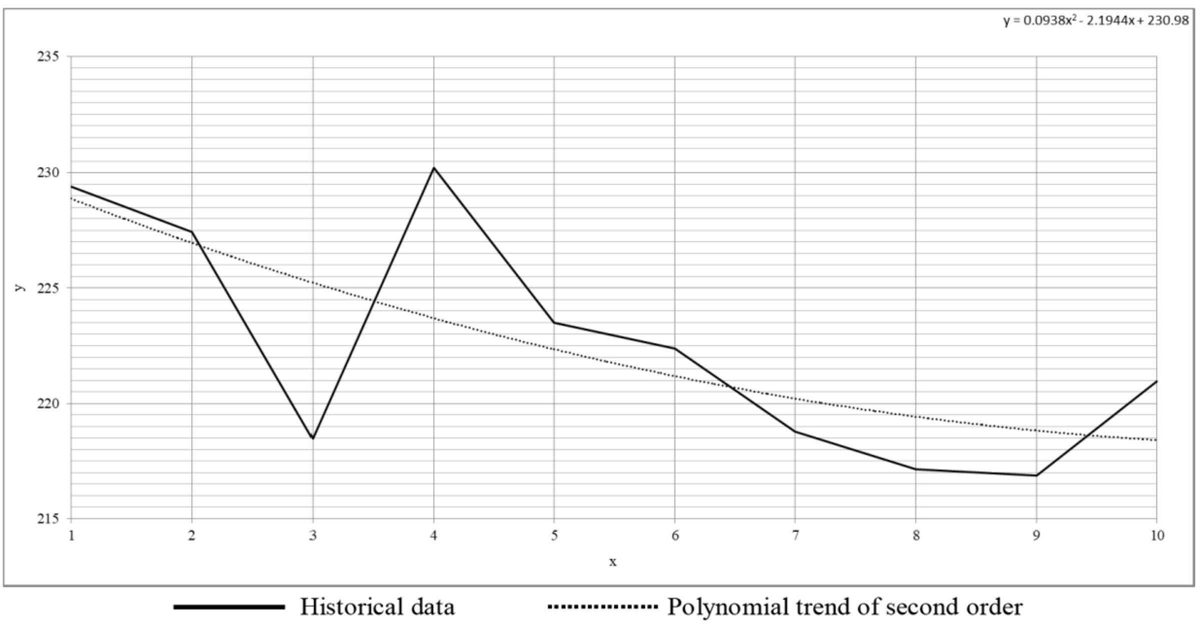

FIGURE 5. 50 km Men's walk

In order to use the equation given in Figure 5, the following is done: the interval of each unit of the "x" axis is 1 and starts with 1. Then each interval is equal to four, because this event is realized each four years and unit 1 must be 1980, because the data is initiated from 1980, and thus ratio between the two variables should be $x = \left(\frac{x_1 - 1976}{4}\right)$.

Now, substituting $x = \left(\frac{x_1 - 1976}{4}\right)$ into equation shown in Figure 5 and the general equation of the road events (50 km Men's walk) is obtained:

$$y = 0.0938\left(\frac{x_1 - 1976}{4}\right)^2 - 2.1944\left(\frac{x_1 - 1976}{4}\right) + 230.98 \tag{5}$$

3. **Results and Discussion.**

This section presents the results obtained from the application of the proposed approach described in the previous section of this paper.

Figure 1 shows the real data and the trend equation for the Men's marathon. The fifth order polynomial equation is the one that fits most. The results show that the fifth order polynomial equations in the years of 1952, 1960, 1964, 1972 to 1988, 2000, 2008 to 2016 are smaller with respect to the real data. The greatest difference is presented in the year of 1968, where the real data are greater with respect to the data obtained by the proposed equation.

Figure 2 shows the real data and the trend equation for the Women's marathon. The fifth order polynomial equation is the one that fits most. The results show that the fifth order polynomial equations in the years of 1988, 1996, 2000, 2012 are smaller with respect to the real data. The greatest difference is presented in the year of 1992, where the real data are greater with respect to the data obtained by the proposed equation.

Figure 3 shows the real data and the trend equation for the 20 km Men's walk. The second order polynomial equation is the one that fits most. The results show that the second order polynomial equations in the years of 1956, 1964, 1972 to 1980, 1988, 1996, 2000, 2012 are smaller with respect to the real data. The greatest difference is presented in the year of 1968, where the real data are greater with respect to the data obtained by the proposed equation.

Figure 4 shows the real data and the trend equation for the 20 km Women's walk. The power equation is the one that fits most. The results show that the power equations in the years of 2000, 2008, 2012 are smaller with respect to the real data. The greatest difference is presented in the year of 2012, where the real data are smaller with respect to the data obtained by the proposed equation.

Figure 5 shows the real data and the trend equation for the 50 km Men's walk. The second order polynomial equation is the one that fits most. The results show that the second order polynomial equations in the years of 1988, 2004 to 2012 are smaller with respect to the real data. The greatest difference is presented in the year of 1988, where the real data are smaller with respect to the data obtained by the proposed equation.

The record times of the athletes for road events in the Olympic Games are: for the Men's marathon is of 126.53 minutes in the year of 2008, for the Women's marathon is of 143.12 minutes in the year of 2012, for the 20 km Men's walk is of 78.77 minutes in the year of 2012, for the 20 km Women's walk is of 85.03 minutes in the year of 2012, for the 50 km Men's walk is of 216.88 minutes in the year of 2012.

Now, substitute $x_1 = 2020$ and 2024 into Equations (1) to (5) to predict the times for the Olympic Games 2020 and 2024 of athletic in the road events (see Table 6).

According to the data presented in Table 6 that predicts the times for the road events, it is observed that the only road event that will break the records in the next Olympic Games is for Women's marathon athletes in the year of 2024.

Table 6. Olympic Games of athletic in the road events for the 2020 and 2024

Road Events	Olympic Games in 2020 Minutes	Olympic Games in 2024 Minutes
Men's marathon	129.73	131.77
Women's marathon	143.93	138.41
20 km Men's walk	78.87	79.14
20 km Women's walk	86.46	86.24
50 km Men's walk	218.19	218.15

4. **Conclusions.** The major contribution of this research is the proposal of the polynomial equation of fifth order for the Men's marathon, the polynomial equation of fifth order for the Women's marathon, the polynomial equation of second order for the 20 km Men's walk, the power equation for the 20 km Women's walk, and the polynomial equation of second order for the 50 km Men's walk. Five equations were created through the real data, and by EXCEL are obtained.

Using the equations to predict the times for the Olympic Games of athletic in the road events in the years of 2020 and 2024 for the Men's marathon must be 129.73 and 131.77 minutes respectively, for the Women's marathon must be 143.93 and 138.41 minutes respectively, for the 20 km Men's walk must be 78.87 and 79.14 minutes respectively, for the 20 km Women's walk must be 86.46 and 86.24 minutes respectively, and for the 50 km Men's walk must be 218.19 and 218.15 minutes respectively.

The accessibility of data in the form of results from Olympic Games, world records and world best performances in a specific year allows the analysis of performances in any number of events. From these analyses, changes in performance over time can be observed and predictions of future performance can be made utilizing the mathematical models.

The suggestions for future research can be to obtain the mathematical equations in another type of sport, either individually or by group.

REFERENCES

[1] Wikipedia, *Olympic Games*, https://en.wikipedia.org/wiki/Olympic_Games.

[2] K. Prendergast, What do world running records tell us?, *Modern Athlete and Coach*, vol.28, pp.33-36, 1990.

[3] I. Heazlewood and G. Lackey, The use of mathematical models to predict elite athletic performance at the Olympic Games, *The 3rd Conference on Mathematics and Computers in Sport*, Bond University, Queensland, pp.185-206, 1996.

[4] T. Heazlewood, Prediction versus reality: The use of mathematical models to predict elite performance in swimming and athletics at the Olympic Games, *Journal of Sports Science and Medicine*, vol.5, no.4, pp.480-487, 2006.

[5] T. Arnold and J. Godbey, Introducing linear regression: An example using basketball statistics, *Journal of Economics and Finance Education*, vol.11, no.2, pp.113-130, 2012.

[6] M. Partington and C. Cushion, An investigation of the practice activities and coaching behaviors of professional top-level youth soccer coaches, *Scandinavian Journal of Medicine & Science in Sports*, vol.23, no.3, pp.374-382, 2013.

[7] G. Ahamad, S. K. Naqvi and M. M. S. Beg, A model for talent identification in cricket based on owa operator, *International Journal of Information Technology & Management Information System*, vol.4, no.2, pp.40-55, 2013.

[8] K. Jackson, Measuring the similarity between players in Australian football, *Proc. of the 13th Australasian Conference on Mathematics and Computers in Sport*, Melbourne, Victoria, Australia, 2016.

[9] M. Huynh and K. Marshall, Training badminton players and knowing when to throw in the tau, *Proc. of the 13th Australasian Conference on Mathematics and Computers in Sport*, Melbourne, Victoria, Australia, pp.63-68, 2016.

[10] L. Devoir, B. Ofoghi, D. Dwyer, C. MacMahond and J. Zeleznikow, Modelling the third version of the cycling omnium, *Proc. of the 13th Australasian Conference on Mathematics and Computers in Sport*, Melbourne, Victoria, Australia, pp.98-103, 2016.
[11] B. O'Bree and I. Grundy, Player ratings in PGA golf, *Proc. of the 13th Australasian Conference on Mathematics and Computers in Sport*, Melbourne, Victoria, Australia, pp.122-127, 2016.
[12] Wikipedia, *Athletics at the Olympic Games*, https://en.wikipedia.org/wiki/Athletics_at_the_Summer_Olympics.