INTERNATIONAL JOURNAL OF RESEARCHES IN SOCIAL SCIENCES AND INFORMATION STUDIES
© VISHWASHANTI MULTIPURPOSE SOCIETY (Global Peace Multipurpose Society) R. No.MH-659/13(N)
www.vmsindia.org

FORECASTING THE PERFORMANCE OF SPRINTERS IN RIO OLYMPICS

Bhanu K.S., Hatewar L.M., Mahurkar M.P. and Kayarwar A.B.

Department of Statistics, Institute of Science, Nagpur (M.S) India

Abstract:

Introduction:

Olympic Games:

The modern Olympic Games or Olympics are the leading international sporting event 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 to be 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 apart.

Over 13,000 athletes compete at the Summer Olympic Games and Winter Olympic Games in 33 different sports and nearly 400 events. The Games have grown so much that nearly every nation is now represented. The first, second, and third-place finishers in each event receive Olympic medals: gold, silver, and bronze, respectively.

In athletics and track and field, sprints (or dashes) are races over short distances. They are among the oldest running competitions. There are three sprinting events which are currently held at the Summer Olympics and outdoor World Championships: the 100 meters, 200 meters, and 400 meters. Olympic middle and long-distance races test the speed, strength and stamina of the competitors in five different events, ranging from 800 meters to the marathon.

Objective:

In this paper, we will attempt to predict the performance (as measured in time) of gold, silver and bronze medal winners in Rio Olympics held in August 2016. The forecasts were made for 100 meter sprint for both men and women.

Methodology:

Data Collection

The data used for our study was collected from the following websites

(i) www.olympic.org/london-2012-summer-olympics

(ii) www.databasesports.com/Olympics and hence data can be said secondary data on the basis of collection. Our data consists of time taken by the sprinters (Men & Women) for Gold, Silver and Bronze medal positions in the respective years. Further, since our data is noted according to the time of occurrence, the resultant series of observations is called a time series. On the other hand, the variable used for the study is taking any value in its range of variat ion, thus it is a continuous variable.

Statistical Methods

Exploratory Data Analysis

In statistics, Exploratory Data Analysis (EDA) is an approach for analyzing data sets to summarize their main characteristics, often with visual methods. Primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. There are a number of tools that are useful for EDA. Some of the typical graphical techniques used in EDA are Box plot, Histogram, Stem-and-leaf plot, Scatter plot. Here, we have used scatter plot to explore the data initially.

Regression Analysis

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables. Regression analysis is widely used for prediction and forecasting.

Poisson Regression Model for Continuous Variables

It is well known that Poisson Regression Model is applied for count data. However, in special cases it may be applied for Continuous Variables also. Generalized linear models (GLM) with the log link function are useful in modeling continuous positive outcomes. The generalised linear model (GLM) generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

Suppose that we have a positive dependent variable Y and a predictor variable X and we want to estimate a model,

$$\log\big(E(Y/X)\big) = X_i' \boldsymbol{\beta}$$

or equivalently,

$$\mathrm{E}(\mathbf{Y}/\mathbf{X}) = \exp(\mathbf{X}_i' \boldsymbol{\beta})$$

$$\mathrm{E}(\mathbf{Y}/\mathbf{X}) = \exp(\beta_0 + \beta_1 \mathbf{X}')$$

where, β0 is the slope and β1 is the intercept.

The above equation is also referred to as a Poisson regression model. It is also sometimes referred to as a log linear model. Poisson regression is a form of regression analysis used to model count data and contingency tables.

But it can be defined for continuous dependent variables. Also, the response variable *Y* does **not** need to be an integer for the estimator based on the Poisson likelihood function to be consistent. The data does not even need to be Poisson. This has been shown by Gourieroux, Monfort and Trognon (1984).

There is also some encouraging simulation evidence from Santos Silva and Tenreyro (2006), where the Poisson comes in best-in-show.

Data analysis and discussion

Exploratory Data Analysis

For Exploratory Data Analysis, we plotted the scatter plots of the data to detect the presence of an outlier, then we plotted the line of best fit on the same graph. Statistical software R.3.2.3 was used for Exploratory Data Analysis and the results obtained are shown in the following figures.

Figure 1. Scatter plots for Male and Female sprinters for Gold Medal

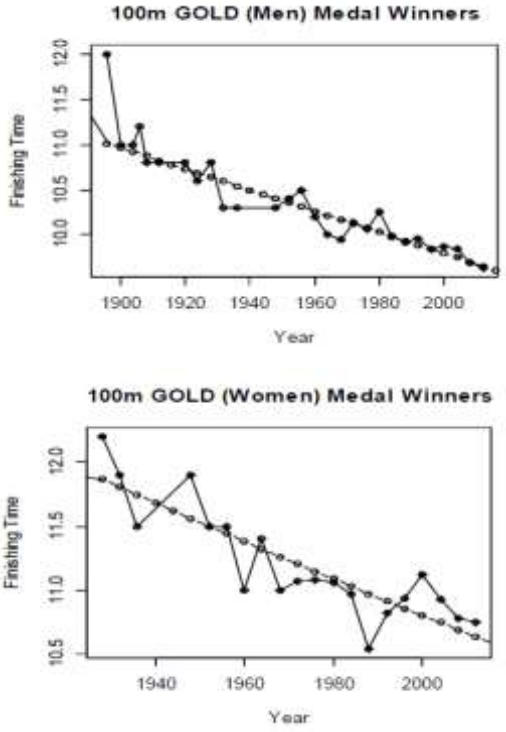
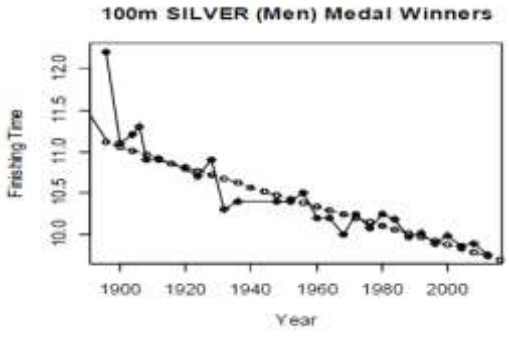


Figure 2. Scatter plots for Male and Female sprinters for Silver Medal



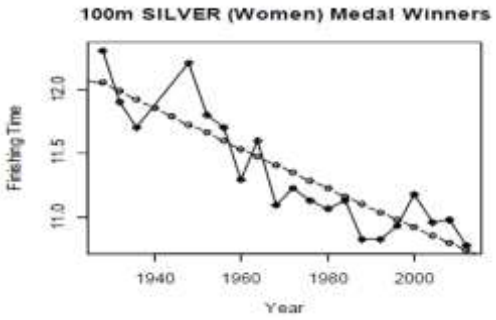
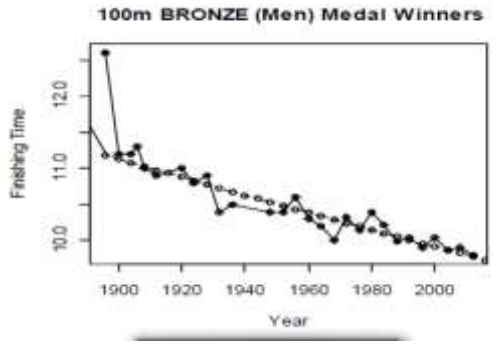


Figure 3. Scatter plots for Male and Female sprinters for Bronze Medal



The solid line connects the individual record data points. A straight line fit to this is shown by the dotted line in figures above.

From the above figures we observe that the general trend in the data is decreasing exponentially. Also, an outlier is present in each of the scatterplot for male sprinters.

Fitting the model

The following values of parameters and deviances are obtained by fitting the Poisson Regression model $(\boldsymbol{Y}\ \boldsymbol{X}\ /\ ) = exp(\boldsymbol{Xi}\ '\ \boldsymbol{\beta})$, for response variable $\boldsymbol{Y}$ (time taken by the sprinters) and regressor $\boldsymbol{X}$ (years) to the data obtained after removing the outliers.

Table1. Summary of the fitted Poisson Regression model for male sprinters

Positions	Intercept	Slope	Residual Deviance	Null Deviance
Gold	4.548210	-0.001133	0.047264	0.484143
Silver	4.550279	-0.001130	0.051257	0.489280
Bronze	4.607612	-0.001157	0.050711	0.512048

Table2. Summary of the fitted Poisson Regression model for female sprinters

Positions	Intercept	Slope	Residual Deviance	Null Deviance
Gold	4.979371	-0.001300	0.080464	0.316632
Silver	5.119256	-0.001364	0.075381	0.363667
Bronze	5.452569	-0.001533	0.060985	0.359754

If Null Deviance is really small it means that the Null model explains the data pretty well likewise with the Residual deviance. So as our observations.

Validation

Following results were obtained after fitting the Poisson Regression Model to infer the 2012 London Olympics timings for Male and Female sprinters.

Table 3. Comparison of Actual and Predicted timings (in sec.) for 2012 London Olympics.

	GOLD		Silver		Bronze	
	Predicted	Actual	Predicted	Actual	Predicted	Actual
100m Men	9.65	9.63	9.75	9.75	9.78	9.79
100m Women	10.64	10.75	10.74	10.78	10.68	10.81

Observing Table 3, we found that the actual and predicted results are very close and some of them exactly coincides, which indicates that the Poisson Regression Model is best fit of the model.

Conclusions:

On the basis of our analysis, we predicted the results for Rio Olympics 2016 which are presented in the following table.

Table 4. Predicted timings (in sec.) for 2016 Rio Olympics

	GOLD	SILVER	BRONZE
100m Men	9.61	9.7	9.73
100m Women	10.58	10.69	10.62

References:

1. Douglas Montgomery: Introduction to linear Regression Analysis
2. Gourieroux, Monfort and Trognon, Pseudo Maximum Likelihood Methods: Application to Poisson Models, Econometrica, Vol. 52, no. 3, 1984
3. Santos, Silva and Tenreyro, The review of Economics and Statistics, November 2006, 88(4): 641-658

*****