Linear regression of 100m World Record

Benjamin Abro

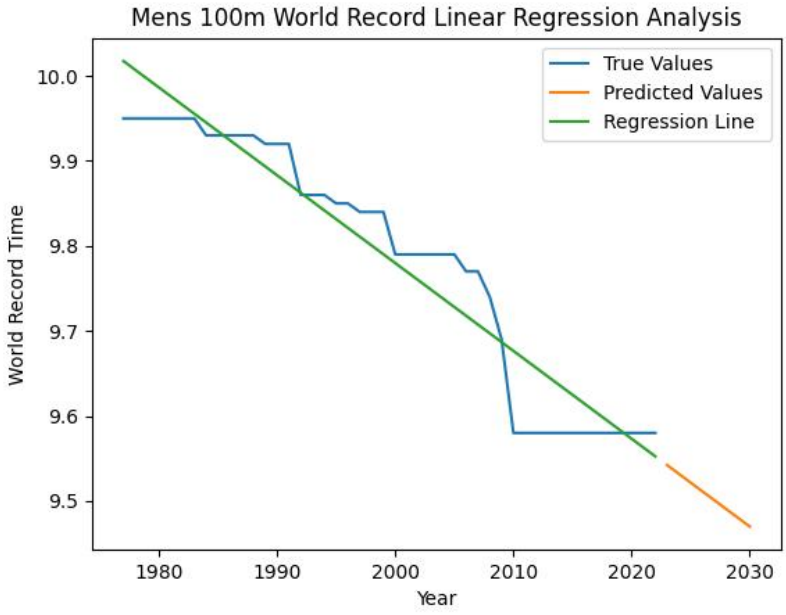
Abstract

We are predicting the times of the 100m record for each year. The data is available at https://en.wikipedia.org/wiki/Men%27s_100_metres_world_record_progression, and the code used is available at https://github.com/benabro/assignment4

Plots

Below are the predicted men's world record 100m running times for 2023 through to 2030, and a plot containing the linear regression predictions, the actual values, and our regression line.

Time	Date
9.542203	2023
9.531871	2024
9.521539	2025
9.511208	2026
9.500876	2027
9.490544	2028
9.480212	2029
9.469880	2030



Analysis

Our regression equation is World Record Time = -0.0103*Year + 30.4434. The R-Squared value corresponding to our linear model is 0.918 which, along with sufficiently small p values, indicates that the model is a good fit. However I do not believe that our model is a good fit in practice. It does not make sense for the 100m world record to decrease linearly as each reduction in the world record time is exponentially harder than the previous reduction. Our model would predict that after the year 2955 the 100m world record time would be negative, this is impossible. Rapid technological change in the 80's and 90's, along with Usain Bolt decreasing the record by an astounding 0.19 seconds in years, have led to the record appearing to be decreasing linearly on the scale that we observed. If we applied certain logistic or exponential transformations to our predictor this may result in a model that would produce better predictions even if the model has a worse fit.