

A COMPARISON BETWEEN STEPWISE REGRESSION MODELS TO PREDICT FOOTBALL GAMES

Natanael Blomberg*

June 2022

Abstract

This study analyzes and compares which variables have the most significant impact on football games. Four models have been created using linear regression and different stepwise regression procedures. The methods used for comparison are backward elimination, forward selection, AIC backward elimination, and AIC forward selection. Additionally, various transformations have been applied, such as a Yeo-Johnson transformation and others, to see if the model could be improved to be more accurate. The model was then used to predict the outcome and analyze which stepwise regression models gave the best prediction. It was discovered that most of the response variables already described the explanatory variable so well that no transformation was needed. Through various prediction tests and measurements, it was concluded that AIC backward elimination and AIC forward selection provided the best models for predicting the outcome of football matches.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: natanaelblomberg@gmail.com. Supervisor: Ola Hössjer & Nils Engler.