

Contributions

1 Introduction

2 Data Description

The data of players in FIFA20 was originally found on Kaggle https://www.kaggle.com/datasets/stefanoleone992/fifa-20-complete-player-dataset?select=players_20.csv, it was scraped from the publicly available website using python script <https://sofifa.com>.

Responsive Variable

The response variable used in this research was “overall”, it is the overall attribute of one player. This variable measures the players overall rating in the game, ranging from 0 to 100.

Table 1: Descriptive Statistic of Response Variable

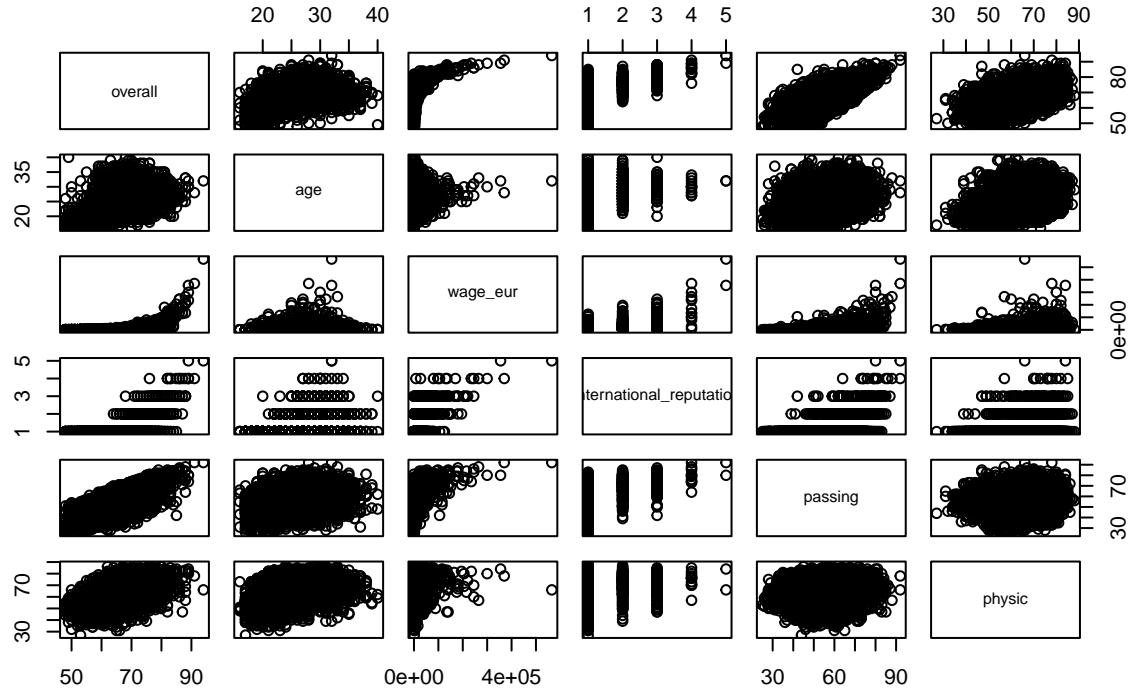
min	max	mean	median	standard_deviation
48	94	66.43309	66	6.819944

The mean of the overall is 66.43

Model Assumptions

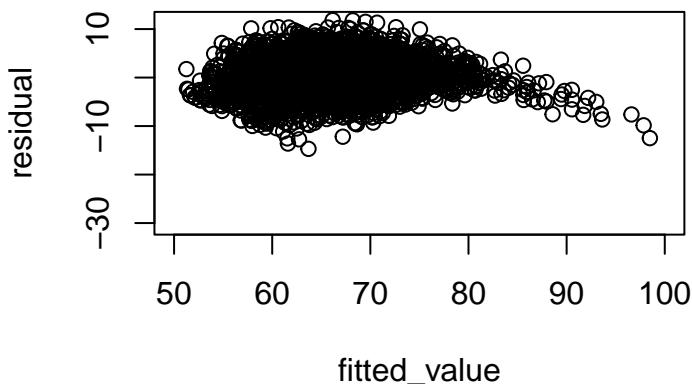
The responsive variable “overall” is the overall rating of one player, we believe the overall rating of a player is largely influenced by their attributes like age, wages, pace score, shooting score and so on. When the player gets higher score in these attribute, their overall rating should also be higher. We think the changes in predictors should result in proportional changes in the response variable.

Response against predictor



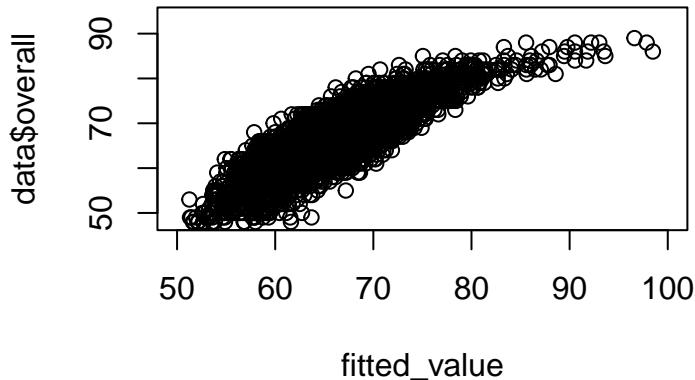
The response variable has a linear relationship with some of the predictor variables. meaning that as the predictors change, the response variable changes in a consistent, straight-line fashion. The mean of each predictor is related to each other predictor in no more complicated way than linearly.

Residual against fitted value

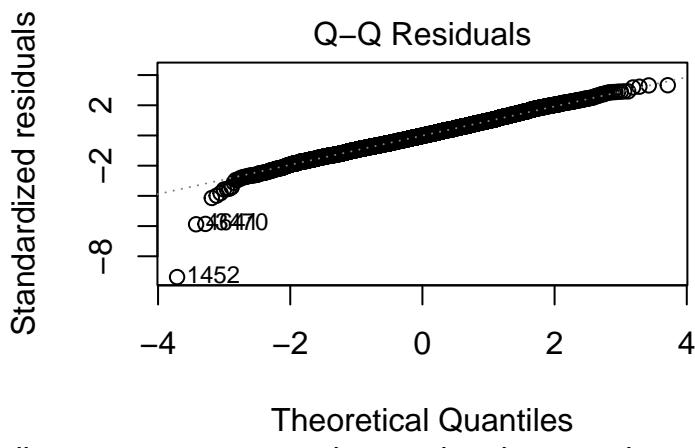


The residual plot has some outlier on the right side, but overall the plot doesn't show a clear pattern. This implies the population satisfies Linearity, constant variance, and uncorrelated errors assumptions.

Response against fitted value



Points closely follow the 45-degree line (i.e., observed \approx predicted), the model is likely performing well. The mean responses are a single function of a linear combination involving coefficients. This satisfies conditional mean response condition.



all ~ age + wage_eur + international_reputation + pass

The most points on the QQ plot are on a straight line, with some outliers on each side. Overall this implies the population satisfies the normal errors assumption.

Predictors

Table 2: Descriptive Statistic of Predictors

predictors	min	max	mean	median	standard_deviation
age	16	40	25.104885	25	4.527917e+00
wage_eur	0	565000	9769.088670	3000	2.238750e+04
international_reputation	1	5	1.104269	1	3.820632e-01
passing	25	92	57.128284	58	1.055442e+01
physic	27	88	64.792693	66	9.850584e+00

Preliminary Results

References