August Decz

Dr. Stuart

STAT 408 - Applied Regression Analysis

5th Dec, 2025

Modeling Player Market Value in FIFA22: A Regression Analysis

Introduction

FIFA 22 is a soccer video game previously owned by Electronic Arts (EA), also known for famous titles such as Madden, NCAA, the Sims, and many more. Personally, I have been playing FIFA since 2012 when I was a kid, and have grown a strong affinity for its playstyle and frequent updates mimicking the current state of the soccer world. Each player in the game is created with many factors (other than graphical and animations) that determine their overall rating (e.g. Messi is a 90 overall but has 87 speed, 94 dribbling etc.), and how those ratings actually transfer to in-game play typically determines market value. The FIFA market is a (mostly) completely open laissez-faire market where supply and demand are primary factors. Higher rated players (or more known players like Messi) are worth more coins than lower rated players. Note that since the game is from over two years ago EA closes the open market down and final market values are unchangeable.

The dataset I found uses only base cards from the version of the game from 2023, we were unable to find a more recent dataset that has the proper columns to answer the questions we strive to answer. Using only base cards eliminates the factor of limited availability (as promo or special cards are harder to find or may have only been available for a select amount of time e.g. Team of the Year cards). The full dataset consists of 16710 observations (each observation is a base player card) and 65 variables describing the player including nationality, height, etc. Note that not all attributes for a player are included in determining the rating, typically statistics such as pace(speed), shooting, passing, dribbling, defense, and physicality are averaged to find a player's rating. However, having played FIFA for a long time I know that taller players are typically valued more (and other examples) so I will be including many different variables in my model to predict player market value.

The goal of this analysis is to build a statistical model that identifies which player attributes best predict market value. By applying regression techniques to a large and well-structured dataset, we aim to understand how FIFA assigns value to different player profiles and how accurately these attributes explain variation in the in-game market.

Data Exploration

The FIFA 22 dataset contains 16,710 player observations and 65 variables of mixed types, including string, categorical, date, and numeric fields. String variables describe player characteristics such as Name, Nationality, and Club. A small number of variables are categorical, including Preferred.Foot (right or left) and Work.Rate (e.g., High/Medium/Low). The majority of variables are numerical player attributes such as Crossing, Finishing, Volleys, Dribbling, and many others, each scored on a 0-100 scale representing skill proficiency. The dataset also includes two date variables, Contract.Valid.Until and Date.Joined, which describe aspects of a player's tenure with their current club. The response variable for this analysis is Value, which initially appears as a string encoded with currency symbols and suffixes (e.g., "€125.5M or K (million and thousand)"). These values represent player market prices in euros and must be converted to numerical format for analysis.

Before modeling, several preprocessing steps were performed to ensure the dataset was suitable for regression analysis. The variables Value, Height, and Weight initially contained non-numeric characters (e.g., "€", "M", "K", "cm", "kg"). These values were cleaned using a mutate() function to strip units and convert each variable into a numeric format (Code 1, Appendix). Because regression requires a numeric and finite response variable, observations with missing, non-finite, or zero values in Value were removed. This ensured that all remaining player records included a valid market value and could be used in the analysis.

To assess the suitability of the response variable for linear regression, preliminary summary statistics and visualizations were examined. As shown in Code 2 (Appendix), the variable Value ranges from $€15,000$ to $€94,000,000$, with a median of $€1,300,000$ and a mean of $€3,491,005$. The large difference between the mean and median, combined with a substantial variance of 6.99×10^{13}, indicates a highly uneven distribution with extreme high-value outliers.

Figure 1 displays a histogram of Value, which shows a pronounced right skew, with most players clustered at relatively low prices and a long tail of high-value players. This skewness suggests that the normality and constant variance assumptions required for linear regression are likely violated when using the raw response variable. As a result, a transformation of Value appears warranted prior to model fitting.

This initial exploratory analysis motivates the use of the Box–Cox procedure in the next section to formally determine an appropriate transformation for stabilizing variance and improving normality.

Model Building

To begin constructing the regression model, a cleaned dataset named fifa_clean was created to isolate variables relevant to predicting player market value. All non-numeric identifiers (e.g., Name, Nationality, Club) were removed, leaving only numerical or categorical attributes that may influence the response variable *Value*. These include Age, Overall rating, Preferred Foot, Skill Moves, Height, Weight, and a wide array of FIFA skill attributes such as Crossing, Finishing, Dribbling, Acceleration, SprintSpeed, Strength, Positioning, Vision, and Defensive Awareness (Code 3, Appendix).

An initial multiple linear regression model was then fit using all selected variables as predictors of Value (Code 4a, Appendix). Diagnostic checks, including the Residuals vs. Fitted plot (Figure 2a) and Q–Q plot (Figure 2b), indicated clear violations of the assumptions of homoscedasticity and normality. These issues suggest that the raw response variable is not suitable for linear modeling in its current form. To formally determine an appropriate transformation, a Box–Cox analysis was performed (Code 5 and

Figure 3). The estimated λ was approximately −0.067, a value close to zero, indicating that a logarithmic transformation of Value is appropriate to stabilize variance and improve normality.

After determining that a transformation was necessary, a new variable Value_bc was created to represent the natural logarithm of Value (Code 6a). The multiple linear regression model was then refitted using this transformed response (Code 6b). Updated model diagnostics, including the model summary (Code 6c), Residuals vs. Fitted plot (Figure 4a), and Q–Q plot (Figure 4b), were examined to assess the effectiveness of the transformation.

The log transformation resulted in substantial improvements to model performance and diagnostic behavior. The F-statistic increased from 326.7 in the raw-value model to 1.024×10^4 in the transformed model, indicating a much stronger overall relationship between player attributes and market value. Similarly, the adjusted R^2 increased from 0.4227 to 0.9584, showing that the transformed model explains a far greater proportion of the variance in player value. The residual plot demonstrates that variance has stabilized, reducing previous concerns of heteroscedasticity, and the Q–Q plot shows that residuals more closely follow a normal distribution. Together, these results indicate that the log transformation produces a far more appropriate and statistically valid model for predicting FIFA 22 player market value.

Variance Inflation Factors (VIFs) were calculated to assess multicollinearity among the explanatory variables (Code 7a). Several predictors exhibited VIF values well above 10, including StandingTackle (29.7), SlidingTackle (26.1), and Dribbling (~18). These values indicate substantial multicollinearity, which is expected given the strong correlations among FIFA skill attributes. Because high multicollinearity can inflate standard errors and reduce coefficient interpretability, a stepwise variable selection procedure (both directions) was used to obtain a more stable model (Code 7b). During this process, several predictors, including Curve, LongPassing, Acceleration, SprintSpeed, Jumping, Strength, Aggression, Interceptions, Vision, and Penalties, were removed because they did not contribute meaningful unique explanatory power beyond what was already captured by strongly correlated attributes. After stepwise reduction, VIF values decreased considerably (Code 7c) , indicating that the final model reduced most of the multicollinearity present in the initial model.

Interpretation of Results

From the initial model of 35 explanatory variables 24 were retained after stepwise selection. Regarding these retained variables used to explain our response variable Value. The final model explains 95.84% of the variability in log-transformed player market value (Adjusted $R^2 = 0.9584$) , indicating an excellent overall fit and providing a reliable foundation for interpreting the effects of each predictor.

(Code 7b): β0 (intercept), $e^{3.05} = 21.273.$ When all predictors are equal to zero (an unrealistic scenario for FIFA players), the model predicts a baseline value of approximately €21,270; however, the intercept has no meaningful real-world interpretation and serves primarily as a mathematical anchor for the model. Holding all other variables constant, age has the strongest negative effect on market value, where each additional year multiplies expected value by $e^{-0.1221235} = 0.885,$ corresponding to an approximate 11.5% decrease. In contrast, Overall rating is the strongest positive predictor, with each additional point multiplying expected value by $e^{0.19851} = 1.2195$, or roughly a 22% increase in value. Skill Moves also meaningfully increases valuation, where each additional star multiplies value by $e^{0.0409283} = 1.0418$, Height has a small positive effect, multiplying value by $e^{0.0030822} = 1.00309$ per centimeter, whereas weight slightly decreases value with $e^{-0.0014547} = 0.99855$ per kilogram. Several technical attributes positively influence value: Finishing($e^{0.0015508} = 1.00155$), Heading Accuracy ($e^{0.0011038} = 1.00110$), Short Passing ($e^{0.0018788} = 1.00188$), Volleys ($e^{0.0017904} = 1.00179$), FK Accuracy ($e^{0.0009870} = 1.00099$), Reactions ($e^{0.0018249} = 1.00183$), Stamina ($e^{0.0008813} = 1.00088$), Positioning ($e^{0.0015243} = 1.00152$), and Standing Tackle ($e^{0.0013730} = 1.00137$). A few predictors show small or weak positive effects, such as Balance ($e^{0.0005144} = 1.00051$) and Defensive Awareness ($e^{0.0005078} = 1.00051$), while Preferred Foot (Right) multiplies value by $e^{0.0072356} = 1.00726$. Several attributes show negative associations after adjusting for correlations with other skills, including Ball Control ($e^{-0.0014725} = 0.99853$), Agility ($e^{-0.0009222} = 0.99908$), Shot Power ($e^{-0.0007072} = 0.99929$), Long Shots ($e^{-0.0011711} = 0.99883$), Composure ($e^{-0.0020889} = 0.99791$), and Sliding Tackle ($e^{-0.0018641} = 0.99814$), each slightly reducing expected market value. Collectively, these multiplicative relationships illustrate how age, overall rating, technical skills, and physical attributes shape player valuation in FIFA once all other variables are held constant.

To quantify uncertainty around each parameter estimate, (Code 8) 95% confidence intervals were computed for all coefficients in the final model. Most predictors exhibited interval ranges that did not include zero, confirming their statistical significance at the 5% level. These include Age, Overall, Skill Moves, Height, Weight, Finishing, Heading Accuracy, Short Passing, Volleys, FK Accuracy, Agility, Reactions, Stamina, Positioning, Composure, Standing Tackle, and Sliding Tackle. In contrast, several predictors had confidence intervals that crossed zero, including Preferred Foot, Crossing, Balance, and Defensive Awareness, indicating that their true effects may be negligible after controlling for other attributes. These results reinforce the stability of the strongest predictors while also highlighting attributes whose contributions to market value are less certain.

Overall, the final log-transformed regression model demonstrates excellent fit, explaining 95.84% of the variability in FIFA22 player market value. This suggests that the combination of player attributes included in the model captures nearly all of the systematic variation in pricing within the game's transfer market. From a video game analysis perspective, such a strong model indicates that EA's valuation system is highly predictable and largely determined by measurable in-game characteristics such as overall rating, technical abilities, and age. This has implications for future modeling of FIFA markets, as developers and analysts could rely on similar attribute-driven frameworks to understand or even forecast player value in new editions of the game. The high fit also suggests that players and traders within FIFA's Ultimate Team market may be responding to consistent underlying patterns rather than random pricing.

Evaluation and Discussion

A major strength of the model is its exceptional explanatory power, with an adjusted r-squared of 0.9584, and its ability to incorporate a large number of player attributes without overfitting. The use of a log transformation and stepwise selection also produced a stable and interpretable set of predictors. However, the model has limitations. Several included variables are highly correlated by design (e.g., passing attributes, dribbling metrics), and although stepwise selection alleviates some multicollinearity, it does not fully eliminate the redundancy among related player skills. In addition, the model assumes linear and additive effects on the log scale, which may overlook complex interactions (such as chemistry, gameplay styles, or meta changes as updates occurred monthly during the games existence). A more flexible modeling approach, such as random forests, might capture nonlinearities more effectively.

One limitation of the analysis is that the dataset comes from FIFA22, a game released in 2021, meaning market dynamics may differ from those of newer editions. Player values in FIFA are also influenced by in-game promos, pack supply, and real-world events, none of which are incorporated into this dataset of base cards. The dataset is large (over 16,000 players), which benefits statistical power but also magnifies the effect of very small coefficient differences and makes nearly all predictors statistically significant. Finally, because the dataset reflects pre-determined game values rather than real economic markets, the findings apply specifically to EA's internal valuation mechanics and may not generalize beyond the game environment such as popularity or real world events (e.g. World Cup or Euro tournaments).

This model highlights the extent to which FIFA's player valuation system is formulaic and attribute-driven, raising interesting implications for the video game economy. In Ultimate Team, market value directly affects player accessibility, competitive balance, and microtransaction spending behavior. Understanding value determinants can help players optimize team building while also revealing how EA structures incentives within the game economy. Additionally, as FIFA esports and competitive play grow, the ability to quantify value drivers may help analysts forecast meta trends or evaluate player cards pre-release. From a research perspective, the project could be improved by incorporating dynamic in-game pricing data, chemistry links, promo effects, or by analyzing year-to-year changes in logic. The process also taught me the importance of careful data preprocessing, model diagnostics, and thoughtful interpretation when working with large applied datasets.

Appendix

Code 1. Mutate String Variables and Filtering

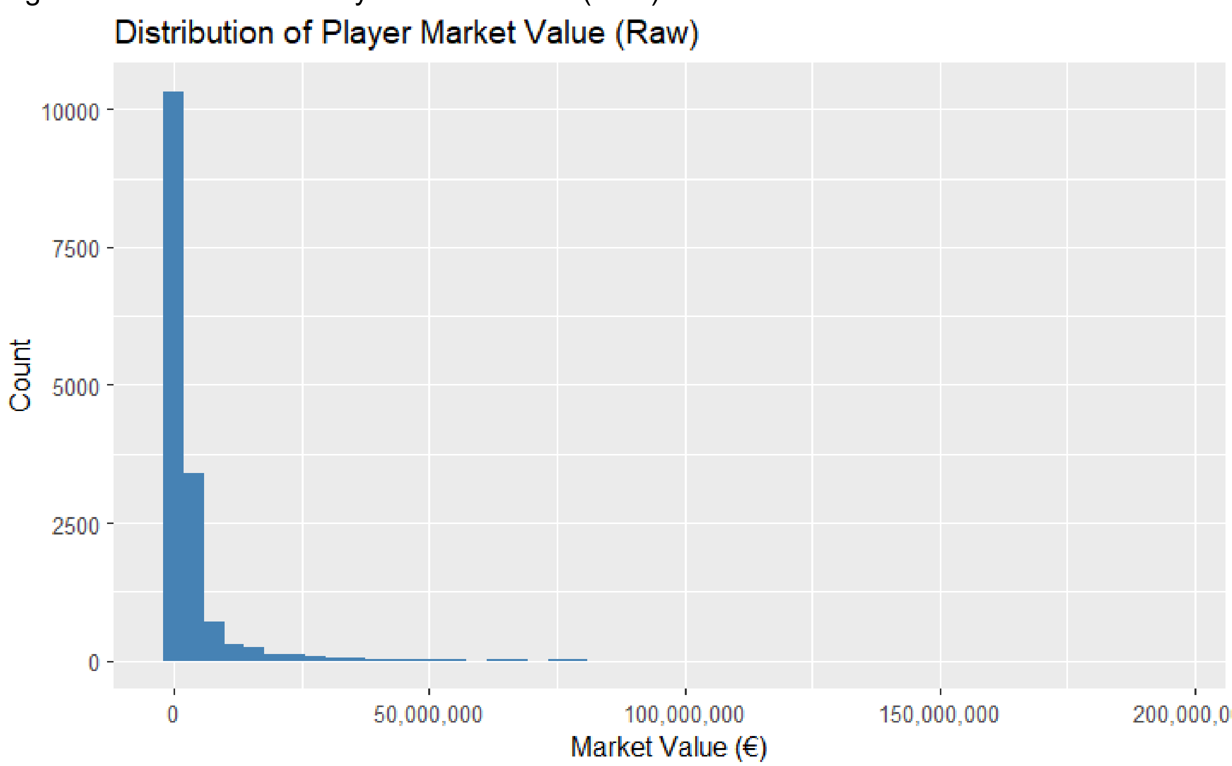
```
mutate(
  Value = gsub("€", "", Value),
  Value = case_when(
    grepl("M", Value) ~ as.numeric(gsub("M", "", Value)) * 1e6,
    grepl("K", Value) ~ as.numeric(gsub("K", "", Value)) * 1e3,
    TRUE ~ as.numeric(Value)
  ),
  Height = as.numeric(gsub("cm", "", Height)),
  Weight = as.numeric(gsub("kg", "", Weight))
) %>%
filter(is.finite(Value), Value > 0) %>%
na.omit()
```

Code 2. Summary Statistics for Response Variable Value

```
{r}
summary(fifa_clean$Value)
var(fifa_clean$Value)
```

```
    Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
   15000    675000   1300000   3491005   2600000 194000000
[1] 6.990145e+13
```

Figure 1. Distribution of Player Market Value (Raw)



Code 3. Selecting Relevant Variables From the FIFA22 Dataset

```r
fifa_clean <- fifa %>%
  dplyr::select(Value, Age, Overall, Preferred.Foot, Skill.Moves, Height, Weight,
         Crossing, Finishing, HeadingAccuracy, ShortPassing, Volleys,
         Dribbling, Curve, FKAccuracy, LongPassing, BallControl,
         Acceleration, SprintSpeed, Agility, Reactions, Balance,
         ShotPower, Jumping, Stamina, Strength, LongShots, Aggression,
         Interceptions, Positioning, Vision, Penalties, Composure,
         StandingTackle, SlidingTackle, DefensiveAwareness) %>%
```

Code 4a. Building an Initial Linear Model

```r
mod_initial <- lm(Value ~ ., fifa_clean)
summary(mod_initial)
```

Code 4b. Initial Linear Model Summary

```
Residuals:
      Min        1Q    Median        3Q       Max
-10744863  -3111361  -1049273   1490104 164712612

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -6.218e+07  2.787e+06 -22.310  < 2e-16 ***
Age                 -5.677e+05  1.474e+04 -38.512  < 2e-16 ***
Overall              8.696e+05  1.912e+04  45.473  < 2e-16 ***
Preferred.FootRight  2.630e+04  1.232e+05   0.214 0.830927
Skill.Moves          3.821e+05  1.182e+05   3.234 0.001224 **
Height               7.941e+04  1.503e+04   5.282 1.30e-07 ***
Weight               1.374e+04  1.286e+04   1.069 0.285100
Crossing            -1.375e+03  7.321e+03  -0.188 0.851003
Finishing            3.266e+04  9.281e+03   3.519 0.000434 ***
HeadingAccuracy     -6.772e+03  6.940e+03  -0.976 0.329194
ShortPassing        -4.518e+03  1.427e+04  -0.317 0.751605
Volleys              3.281e+04  7.906e+03   4.149 3.35e-05 ***
Dribbling           -3.700e+04  1.221e+04  -3.030 0.002449 **
Curve                1.112e+04  7.737e+03   1.437 0.150630
FKAccuracy           9.082e+03  6.514e+03   1.394 0.163240
```
\

```
FKAccuracy            9.082e+03  6.514e+03   1.394 0.163240    
LongPassing           3.586e+03  9.801e+03   0.366 0.714447    
BallControl          -5.510e+04  1.479e+04  -3.726 0.000195 ***
Acceleration         -2.270e+04  1.090e+04  -2.083 0.037254 *  
SprintSpeed           2.398e+04  9.651e+03   2.484 0.012985 *  
Agility              -1.342e+04  8.320e+03  -1.613 0.106855    
Reactions             7.753e+04  1.328e+04   5.840 5.33e-09 ***
Balance               5.419e+04  7.838e+03   6.913 4.92e-12 ***
ShotPower            -3.991e+04  8.537e+03  -4.675 2.96e-06 ***
Jumping              -1.678e+02  5.331e+03  -0.031 0.974894    
Stamina              -1.904e+04  6.085e+03  -3.129 0.001758 ** 
Strength             -3.467e+03  7.692e+03  -0.451 0.652243    
LongShots            -2.996e+04  8.855e+03  -3.383 0.000718 ***
Aggression            1.362e+03  5.937e+03   0.229 0.818488    
Interceptions         5.142e+03  9.282e+03   0.554 0.579583    
Positioning           1.501e+04  9.147e+03   1.641 0.100897    
Vision                3.336e+04  8.245e+03   4.046 5.24e-05 ***
Penalties             9.427e+03  7.307e+03   1.290 0.197021    
Composure             1.354e+04  9.158e+03   1.478 0.139321    
StandingTackle        1.459e+04  1.298e+04   1.124 0.260871    
SlidingTackle        -1.404e+04  1.233e+04  -1.138 0.255116    
DefensiveAwareness    8.020e+03  8.351e+03   0.960 0.336873    
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6353000 on 15535 degrees of freedom
Multiple R-squared:  0.424,	Adjusted R-squared:  0.4227 
F-statistic: 326.7 on 35 and 15535 DF,  p-value: < 2.2e-16
```

Figure 2a. Residual Plot of Initial Linear Model

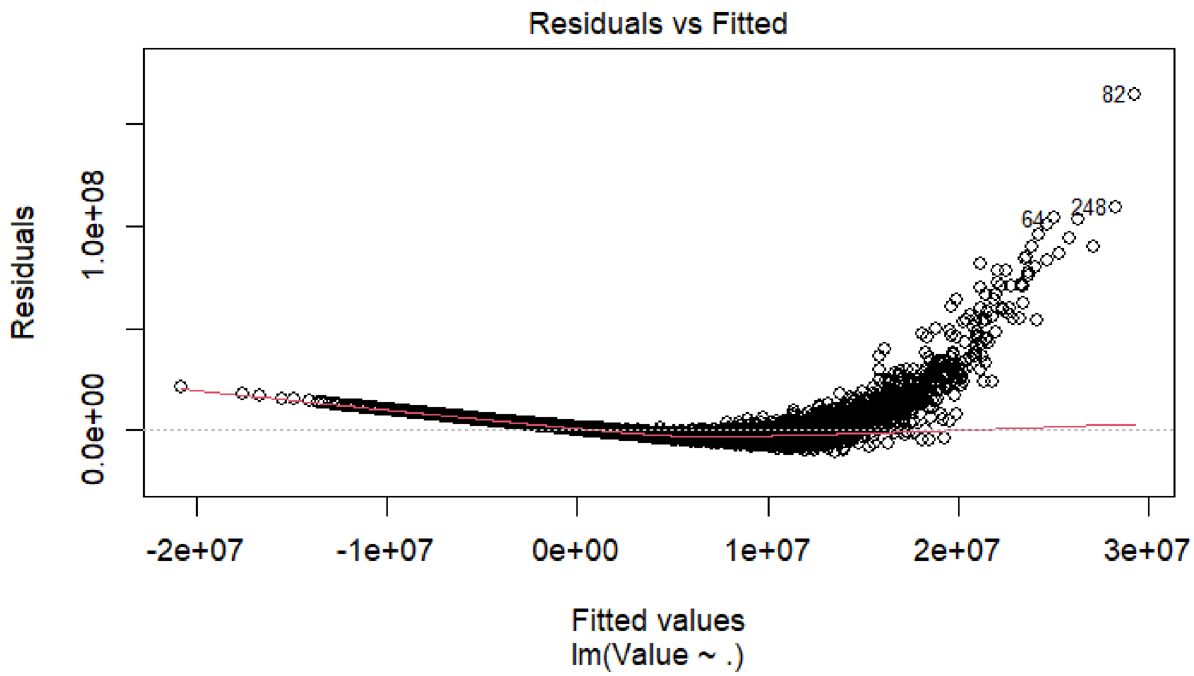
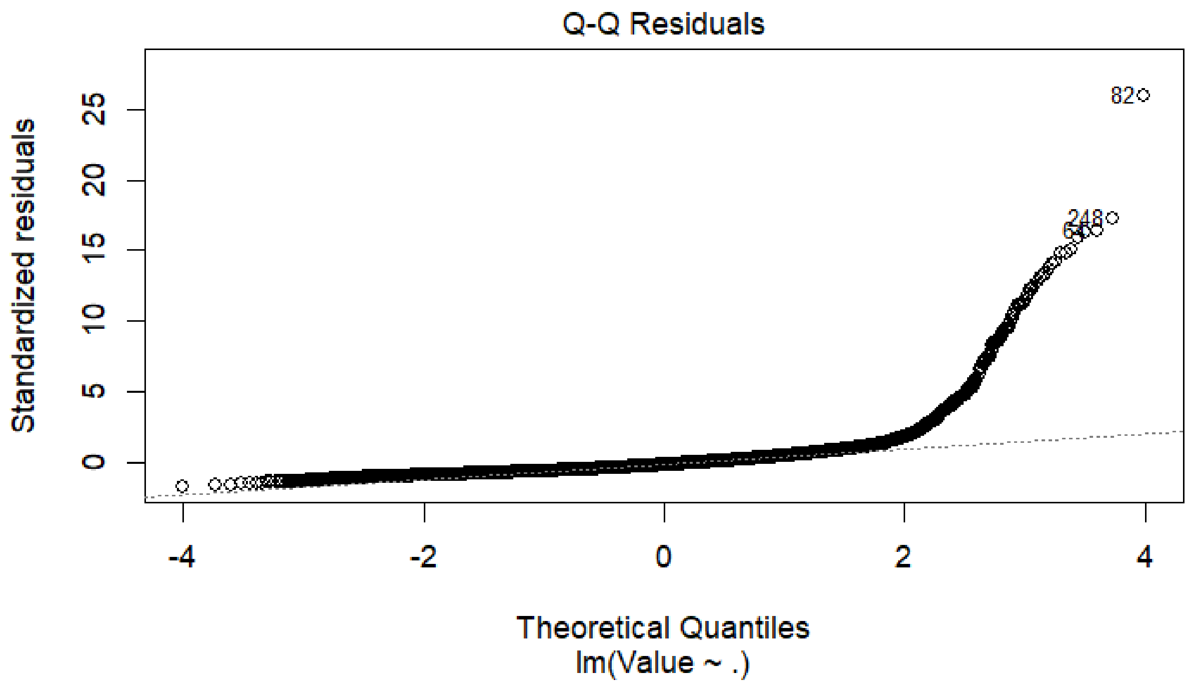


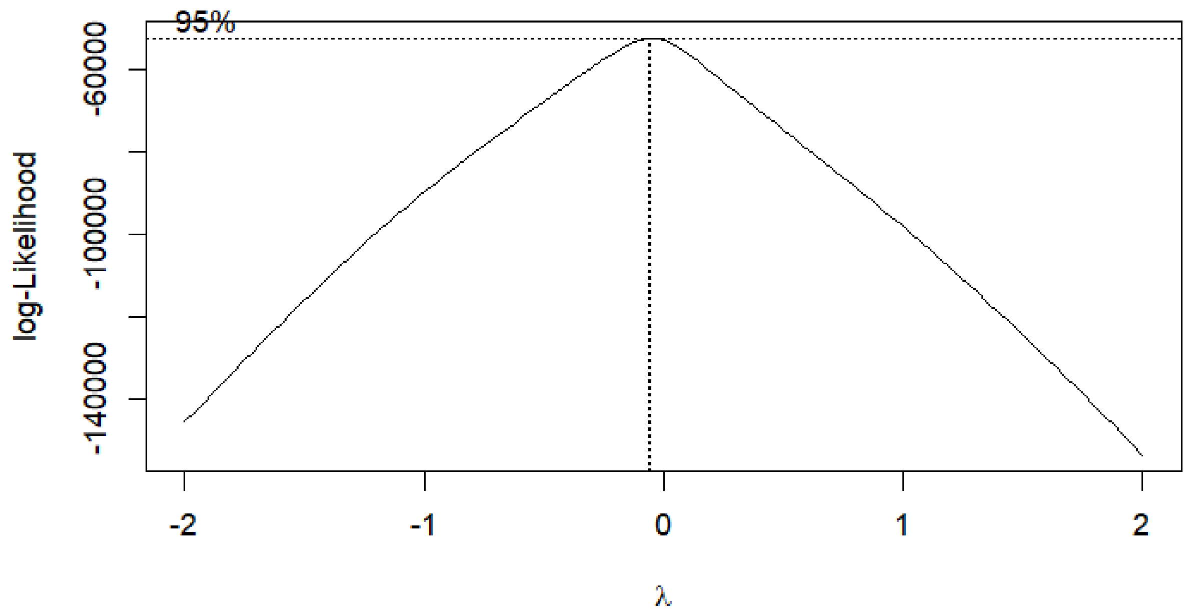
Figure 2b. Normality Plot of Initial Linear Model



Code 5. Box-Cox Analysis of the Initial Linear Model

```{r}
library(MASS)
bc <- boxcox(mod_initial)
lambda <- bc$x[which.max(bc$y)]
lambda
```

Figure 3. Box-Cox Plot of Initial Linear Model



Code 6a. Log transformation of Response Variable

```
fifa_clean <- fifa_clean %>%
  mutate(Value_bc = log(Value))
```

Code 6b. Model Refitting

```
mod_bc <- lm(Value_bc ~ ., data = fifa_clean %>% dplyr::select(-Value))
```

Code 6c. Transformed Model Summary

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.47500 -0.13383  0.00641  0.14795  1.10243
```

```
Coefficients:
                      Estimate Std. Error  t value Pr(>|t|)    
(Intercept)          3.056e+00  1.063e-01   28.764  < 2e-16 ***
Age                 -1.222e-01  5.620e-04 -217.426  < 2e-16 ***
Overall              1.984e-01  7.290e-04  272.151  < 2e-16 ***
Preferred.FootRight  7.300e-03  4.697e-03    1.554 0.120133    
Skill.Moves          4.110e-02  4.505e-03    9.122  < 2e-16 ***
Height               3.075e-03  5.731e-04    5.366 8.19e-08 ***
Weight              -1.598e-03  4.901e-04   -3.262 0.001110 ** 
Crossing            -3.436e-04  2.791e-04   -1.231 0.218193    
Finishing            1.531e-03  3.538e-04    4.327 1.52e-05 ***
HeadingAccuracy      8.409e-04  2.646e-04    3.178 0.001484 ** 
ShortPassing         2.188e-03  5.442e-04    4.020 5.85e-05 ***
Volleys              1.718e-03  3.014e-04    5.701 1.21e-08 ***
Dribbling           -5.119e-04  4.655e-04   -1.100 0.271508    
Curve                1.431e-04  2.950e-04    0.485 0.627621    
FKAccuracy           9.474e-04  2.483e-04    3.815 0.000137 ***
LongPassing         -2.602e-04  3.737e-04   -0.696 0.486148    
BallControl         -1.253e-03  5.637e-04   -2.222 0.026295 *  
Acceleration         1.609e-04  4.154e-04    0.387 0.698482    
SprintSpeed          2.320e-04  3.679e-04    0.630 0.528389    
Agility             -1.096e-03  3.172e-04   -3.456 0.000549 ***
Reactions            1.846e-03  5.061e-04    3.648 0.000265 ***
Balance              4.434e-04  2.988e-04    1.484 0.137851    
ShotPower           -7.779e-04  3.255e-04   -2.390 0.016857 *  
Jumping              1.127e-04  2.032e-04    0.554 0.579330    
Stamina              7.312e-04  2.320e-04    3.152 0.001625 ** 
Strength             1.052e-04  2.932e-04    0.359 0.719728    
LongShots           -1.106e-03  3.376e-04   -3.277 0.001050 ** 
Aggression           2.529e-04  2.263e-04    1.117 0.263824    
Interceptions       -4.905e-05  3.538e-04   -0.139 0.889759    
Positioning          1.512e-03  3.487e-04    4.337 1.45e-05 ***
Vision               6.469e-05  3.143e-04    0.206 0.836925    
Penalties            2.578e-04  2.786e-04    0.925 0.354855    
Composure           -2.096e-03  3.491e-04   -6.004 1.97e-09 ***
StandingTackle       1.388e-03  4.948e-04    2.804 0.005052 ** 
SlidingTackle       -1.851e-03  4.702e-04   -3.937 8.30e-05 ***
DefensiveAwareness   5.140e-04  3.184e-04    1.615 0.106398    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2422 on 15535 degrees of freedom
Multiple R-squared:  0.9584,	Adjusted R-squared:  0.9584 
F-statistic: 1.024e+04 on 35 and 15535 DF,  p-value: < 2.2e-16
```

Figure 4a. Residual Plot of Transformed Model

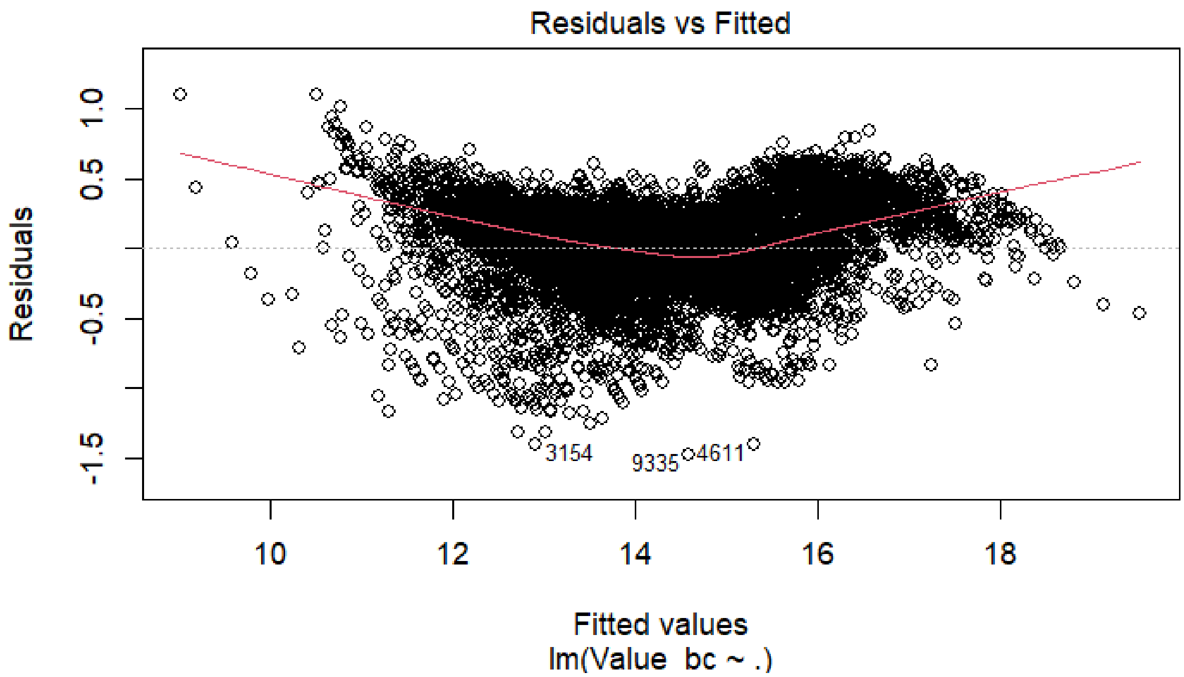
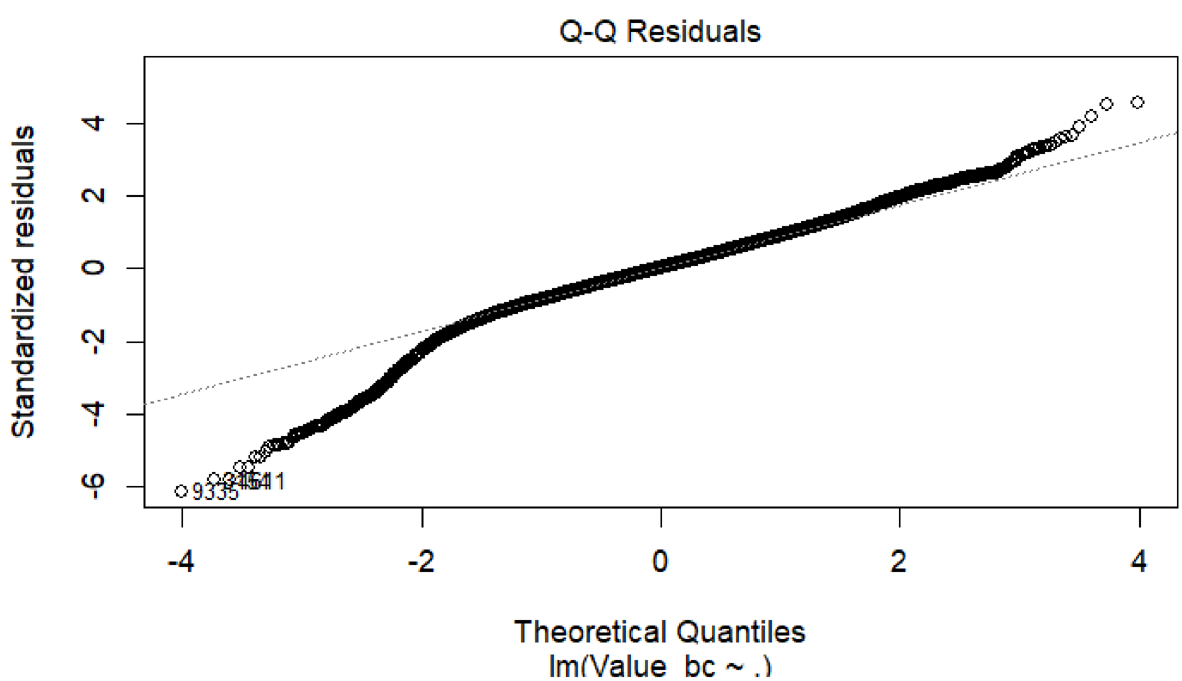


Figure 4b. Normality Plot of Transformed Model



Code 7a. Check for Multicollinearity of the Log Transformed Model

```
                Age             Overall      Preferred.Foot         Skill.Moves
Height                 Weight                Crossing             Finishing
           1.998108            5.794582            1.087650            3.333076
4.129438               3.254866             6.466285             12.473149
   HeadingAccuracy        ShortPassing             Volleys           Dribbling
Curve             FKAccuracy         LongPassing            BallControl
           5.324311           14.502852            7.526063           18.385173
7.563848               4.949327             7.688330             20.561513
      Acceleration         SprintSpeed             Agility           Reactions
Balance                ShotPower           Jumping              Stamina
          10.156489            7.681876            5.591508            4.962193
4.840353               4.528045            1.657447             3.448337
           Strength           LongShots          Aggression       Interceptions
Positioning              Vision             Penalties             Composure
           3.650056           10.933974            3.850211           14.377313
11.826492               4.609198             4.970750             4.152776
    StandingTackle        SlidingTackle DefensiveAwareness
          29.710123           26.099939           10.931198
```

Code 7b. Stepwise/Final Model Summary

```
Call:
lm(formula = Value_bc ~ Age + Overall + Preferred.Foot + Skill.Moves +
    Height + Weight + Crossing + Finishing + HeadingAccuracy +
    ShortPassing + Volleys + FKAccuracy + BallControl + Agility +
    Reactions + Balance + ShotPower + Stamina + LongShots + Positioning +
    Composure + StandingTackle + SlidingTackle + DefensiveAwareness,
    data = fifa_clean %>% dplyr::select(-Value))

Residuals:
    Min      1Q  Median      3Q     Max
-1.4772 -0.1340  0.0067  0.1476  1.1011

Coefficients:
                     Estimate Std. Error  t value Pr(>|t|)
(Intercept)         3.0570830  0.1031621   29.634  < 2e-16 ***
Age                -0.1221235  0.0005161 -236.639  < 2e-16 ***
Overall             0.1985100  0.0007157  277.352  < 2e-16 ***
Preferred.FootRight 0.0072356  0.0046774    1.547 0.121904
Skill.Moves         0.0409283  0.0044335    9.232  < 2e-16 ***
Height              0.0030822  0.0005621    5.483 4.25e-08 ***
Weight             -0.0014547  0.0004541   -3.203 0.001362 **
Crossing           -0.0003816  0.0002571   -1.484 0.137748
Finishing           0.0015508  0.0003446    4.501 6.82e-06 ***
HeadingAccuracy     0.0011038  0.0002127    5.189 2.14e-07 ***
ShortPassing        0.0018788  0.0004428    4.243 2.22e-05 ***
Volleys             0.0017904  0.0002924    6.123 9.41e-10 ***
FKAccuracy          0.0009870  0.0002184    4.518 6.28e-06 ***
BallControl        -0.0014725  0.0004905   -3.002 0.002684 **
Agility            -0.0009222  0.0002756   -3.346 0.000823 ***
Reactions           0.0018249  0.0004980    3.664 0.000249 ***
Balance             0.0005144  0.0002923    1.760 0.078412 .
ShotPower          -0.0007072  0.0003188   -2.219 0.026529 *
Stamina             0.0008813  0.0002153    4.093 4.28e-05 ***
LongShots          -0.0011711  0.0003341   -3.505 0.000458 ***
Positioning         0.0015243  0.0003291    4.632 3.66e-06 ***
Composure          -0.0020889  0.0003420   -6.109 1.03e-09 ***
StandingTackle      0.0013730  0.0004661    2.946 0.003227 **
SlidingTackle      -0.0018641  0.0004619   -4.035 5.48e-05 ***
DefensiveAwareness  0.0005078  0.0003032    1.675 0.094019 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2421 on 15546 degrees of freedom
Multiple R-squared:  0.9584,    Adjusted R-squared:  0.9584
F-statistic: 1.493e+04 on 24 and 15546 DF,  p-value: < 2.2e-16
```
]

Code 7c. Multicollinearity Check After Running Stepwise Selection

```
{r}
vif(step_mod)

                Age             Overall       Preferred.Foot          Skill.Moves
Height              Weight              Crossing             Finishing
           1.685412            5.586770            1.079106            3.229073
3.974015            2.795889            5.489517           11.831106
    HeadingAccuracy         ShortPassing             Volleys           FKAccuracy
BallControl             Agility           Reactions             Balance
           3.442319            9.604312            7.085366            3.830997
15.569713            4.223318            4.806970            4.632247
          ShotPower             Stamina           LongShots          Positioning
Composure       StandingTackle        SlidingTackle DefensiveAwareness
           4.344951            2.970892           10.712908           10.536944
3.985156           26.367107           25.194541            9.918156
```

Code 8. Confidence Interval of the Final Model

```
{r}
confint(step_mod)

                           2.5 %        97.5 %
(Intercept)          2.854873e+00  3.259293e+00
Age                 -1.231351e-01 -1.211120e-01
Overall              1.971071e-01  1.999129e-01
Preferred.FootRight -1.932729e-03  1.640395e-02
Skill.Moves          3.223813e-02  4.961854e-02
Height               1.980382e-03  4.184099e-03
Weight              -2.344877e-03 -5.645372e-04
Crossing            -8.855725e-04  1.223321e-04
Finishing            8.754398e-04  2.226177e-03
HeadingAccuracy      6.868690e-04  1.520742e-03
ShortPassing         1.010926e-03  2.746722e-03
Volleys              1.217220e-03  2.363496e-03
FKAccuracy           5.588056e-04  1.415165e-03
BallControl         -2.433872e-03 -5.111545e-04
Agility             -1.462493e-03 -3.819316e-04
Reactions            8.487037e-04  2.801155e-03
Balance             -5.845981e-05  1.087287e-03
ShotPower           -1.332020e-03 -8.239032e-05
Stamina              4.592441e-04  1.303296e-03
LongShots           -1.825964e-03 -5.161630e-04
Positioning          8.792036e-04  2.169359e-03
Composure           -2.759238e-03 -1.418639e-03
StandingTackle       4.593589e-04  2.286567e-03
SlidingTackle       -2.769605e-03 -9.586838e-04
DefensiveAwareness  -8.654894e-05  1.102080e-03
```