

# Regression Variables Selector

This document shows a step-by-step procedure of how the variables were selected for modeling

## Loading the packages

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.0      v purrr  0.3.4
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lasso2)
```

```
## R Package to solve regression problems while imposing
##   an L1 constraint on the parameters. Based on S-plus Release 2.1
## Copyright (C) 1998, 1999
## Justin Lokhorst <jlokhors@stats.adelaide.edu.au>
## Berwin A. Turlach <bturlach@stats.adelaide.edu.au>
## Bill Venables <wvenable@stats.adelaide.edu.au>
##
## Copyright (C) 2002
## Martin Maechler <maechler@stat.math.ethz.ch>
```

## Loading the datasets

```
s17_1 <- read.csv("~/DSI-SRP1/season2017.csv", encoding = "UTF-8")
s17_2 <- read.csv("~/DSI-SRP1/FPL_2016_17_new.csv")
s18_1 <- read.csv("~/DSI-SRP1/season2018.csv", encoding = "UTF-8")
s18_2 <- read.csv("~/DSI-SRP1/FPL_2017_18_new.csv")
s19_1 <- read.csv("~/DSI-SRP1/season2019.csv", encoding = "UTF-8")
s19_2 <- read.csv("~/DSI-SRP1/FPL_2018_19_new.csv")
```

## Variable selector

```
var_sel <- function(df) {  
  df %>%  
    select(goals_scored,assists,total_points,minutes.played,bonus,  
           bps,goals_conceded,clean_sheets,ict_index,position_index,  
           selected_by_percent)  
}
```

This is done using the seasonal datasets that is going to be used for modeling

```
s17_1n <- var_sel(s17_1)  
s18_1n <- var_sel(s18_1)  
s19_1n <- var_sel(s19_1)  
s_combined <- rbind(s17_1n, s18_1n, s19_1n)  
lasso.s <- l1ce(total_points ~., data = s_combined)  
summary(lasso.s)$coefficients
```

##	Value	Std. Error	Z score	Pr(> Z )
## (Intercept)	16.46621655	1.438894135	11.443661	0.000000e+00
## goals_scored	0.00000000	0.313436323	0.000000	1.000000e+00
## assists	0.00000000	0.325963467	0.000000	1.000000e+00
## minutes.played	0.00000000	0.003115412	0.000000	1.000000e+00
## bonus	0.65961380	0.187782364	3.512650	4.436607e-04
## bps	0.08088562	0.010440822	7.747055	9.325873e-15
## goals_conceded	0.00000000	0.104186649	0.000000	1.000000e+00
## clean_sheets	0.67163353	0.346257177	1.939696	5.241669e-02
## ict_index	0.16194721	0.024962065	6.487733	8.713763e-11
## position_index	0.00000000	0.445673259	0.000000	1.000000e+00
## selected_by_percent	0.00000000	0.139350609	0.000000	1.000000e+00

```
lasso.s <- l1ce(total_points ~ ict_index + bps + clean_sheets + bonus,  
               data = s_combined)  
summary(lasso.s)$coefficients
```

##	Value	Std. Error	Z score	Pr(> Z )
## (Intercept)	25.47307218	0.814252771	31.283986	0.000000e+00
## ict_index	0.12862290	0.017572139	7.319707	2.484679e-13
## bps	0.07502564	0.007530691	9.962651	0.000000e+00
## clean_sheets	0.00000000	0.334994853	0.000000	1.000000e+00
## bonus	0.03378626	0.204580419	0.165149	8.688267e-01