Nalan Akın

0078488

Ball Possession and Goal Expectation

My hypothesis is that in a football match, a team's percentage of ball possession (β_2) predicts the number of expected goals (xG). I usually look at these two stats after a match and doing better in one often predicts the performance in the other one. The game Crystal Palace (60%, 1.48) vs Southampton (40%, 0.72) played on Dec 15th is an example. However, there are opposite cases as well such as the match between Aston Villa (41%, 2.18 xG) vs Leicester (59%, 0.88 xG) that was played on Dec 5th. As a control variable for this relationship, average points per game collected by the team (β_1) as an indicator of overall performance of the team will be included as well. The data was not scraped but all three variables were taken from Premier League 2018-2019 Matches dataset that was prepared by FootyStats.org. There was no missing data.

Firstly, the results obtained through using the function I wrote shows that coefficient of β_2 is 0.022 (standardized: .780 according to SPSS) without the constant and it is a significant predictor of expected goals. The coefficient of β_1, 0.189 (standardized: .195), is also significant.

Then, I used the OLS function of the statsmodels.api module and SPSS to see if their results match with what my functions produced. The coefficients were the same in all of them, all variables were the same with SPSS. The only difference was standardized coefficients. I calculate it by BB = (np.std(X) / np.std(y)) * B but could not find how SPSS calculates it to compare and understand why they differ.

```
In [1]: import xG
coefficient: [0.18905889 0.02230146]
standard error: [0.02623601 0.00077326]
confidence intervals: (array([0.1376363 , 0.02078588]), array([0.24048148, 0.02381704]))
```

```
x1    0.189059
x2    0.022301
dtype: float64
x1     7.210841
x2    28.860051
dtype: float64
                            Test for Constraints
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
c0             0.1891      0.026      7.211      0.000       0.138       0.241
==============================================================================
```

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	ppg	.189	.026	.195	7.211	<.001	.138	.241
	ballpossession	.022	.001	.780	28.860	<.001	.021	.024

a. Dependent Variable: xG

b. Linear Regression through the Origin