2 Pythagorean expectation and the IPL

August 16, 2022

1 Pythagorean Expectation and the Indian Premier League

The Indian Premier League (IPL) is the biggest cricket competition in the world, which has all of the world's best players in an eight week tournament involving eight teams playing sixty games in total. Each team plays every other team, once at home and then away, and the competition finishes with the four best teams competing in semi-finals and then a final.

Cricket, like baseball, is a bat and ball game, where teams score runs and the team scoring the highest number of runs is the winner. There are, of course, many differences, but statistically speaking, we can generate the same Pythagorean statistic that we generated for baseball. Our data here is derived from the competition that took place in 2018.

The IPL is played in the T20 format, in which each team has up to 120 balls to score as many runs as they can (the game takes less than three hours to complete). One difference from baseball is that runs are much easier to score - in the IPL an average score is 170 runs - and outs (wickets) are much more costly - each team has only ten outs(called wickets) in the entire game, and if you run out of wickets before the 120 balls have been bowled (pitched) then your inning is over.

With this background, let's construct the Pythagorean Expectation for the IPL in 2018.

In [19]:
```
# As with the previous notebook, we first important the packages we will need to proc

import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt
import seaborn as sns
```

In [20]:
```
# Now we import the data, which comes in the form of a list of games played in the 20
# We print out the list of variables names in the dataframe

IPL18 = pd.read_excel('../../Data/Week 1/IPL2018teams.xlsx')
print(IPL18.columns.tolist())
```

```
['scorecard_id', 'start_date', 'phase', 'name', 'home_team', 'away_team', 'toss_winner', 'toss_
```

In [21]:
```
# We can see what our dataframe looks like simply by typing its name:

IPL18
```

Out[21]:

	scorecard_id	start_date	phase	\
0	1056637	2018-04-07	NaN	
1	1056638	2018-04-08	NaN	
2	1056639	2018-04-08	NaN	
3	1056640	2018-04-09	NaN	
4	1056641	2018-04-10	NaN	
5	1056642	2018-04-11	NaN	
6	1056643	2018-04-12	NaN	
7	1056644	2018-04-13	NaN	
8	1056645	2018-04-14	NaN	
9	1056646	2018-04-14	NaN	
10	1056647	2018-04-15	NaN	
11	1056648	2018-04-15	NaN	
12	1056649	2018-04-16	NaN	
13	1056650	2018-04-17	NaN	
14	1056651	2018-04-18	NaN	
15	1056652	2018-04-19	NaN	
16	1056653	2018-04-20	NaN	
17	1056654	2018-04-21	NaN	
18	1056655	2018-04-21	NaN	
19	1056656	2018-04-22	NaN	
20	1056657	2018-04-22	NaN	
21	1056658	2018-04-23	NaN	
22	1056659	2018-04-24	NaN	
23	1056660	2018-04-25	NaN	
24	1056661	2018-04-26	NaN	
25	1056662	2018-04-27	NaN	
26	1056663	2018-04-28	NaN	
27	1056664	2018-04-29	NaN	
28	1056665	2018-04-29	NaN	
29	1056666	2018-04-30	NaN	
30	1056667	2018-05-01	NaN	
31	1056668	2018-05-02	NaN	
32	1056669	2018-05-03	NaN	
33	1056670	2018-05-04	NaN	
34	1056671	2018-05-05	NaN	
35	1056672	2018-05-05	NaN	
36	1056673	2018-05-06	NaN	
37	1056674	2018-05-06	NaN	
38	1056675	2018-05-07	NaN	
39	1056676	2018-05-08	NaN	
40	1056677	2018-05-09	NaN	
41	1056678	2018-05-10	NaN	
42	1056679	2018-05-11	NaN	
43	1056680	2018-05-12	NaN	
44	1056681	2018-05-12	NaN	
45	1056682	2018-05-13	NaN	
46	1056683	2018-05-13	NaN	

```
47       1056684 2018-05-14                    NaN
48       1056685 2018-05-15                    NaN
49       1056686 2018-05-16                    NaN
50       1056687 2018-05-17                    NaN
51       1056688 2018-05-18                    NaN
52       1056689 2018-05-19                    NaN
53       1056690 2018-05-19                    NaN
54       1056691 2018-05-20                    NaN
55       1056692 2018-05-20                    NaN
56       1056693 2018-05-22  1st Preliminary Final
57       1056694 2018-05-23  2nd Preliminary Final
58       1056695 2018-05-25  3rd Preliminary Final
59       1056696 2018-05-27                  Final

                                                 name  \
0                            Wankhede Stadium, Mumbai
1          Punjab Cricket Association Stadium, Mohali
2                               Eden Gardens, Kolkata
3   Rajiv Gandhi International Stadium, Uppal, Hyd...
4            MA Chidambaram Stadium, Chepauk, Chennai
5                      Sawai Mansingh Stadium, Jaipur
6   Rajiv Gandhi International Stadium, Uppal, Hyd...
7                    M Chinnaswamy Stadium, Bangalore
8                            Wankhede Stadium, Mumbai
9                               Eden Gardens, Kolkata
10                   M Chinnaswamy Stadium, Bangalore
11         Punjab Cricket Association Stadium, Mohali
12                              Eden Gardens, Kolkata
13                           Wankhede Stadium, Mumbai
14                     Sawai Mansingh Stadium, Jaipur
15         Punjab Cricket Association Stadium, Mohali
16   Maharashtra Cricket Association Stadium, Gahunje
17                              Eden Gardens, Kolkata
18                   M Chinnaswamy Stadium, Bangalore
19  Rajiv Gandhi International Stadium, Uppal, Hyd...
20                     Sawai Mansingh Stadium, Jaipur
21                             Feroz Shah Kotla, Delhi
22                           Wankhede Stadium, Mumbai
23                   M Chinnaswamy Stadium, Bangalore
24  Rajiv Gandhi International Stadium, Uppal, Hyd...
25                             Feroz Shah Kotla, Delhi
26   Maharashtra Cricket Association Stadium, Gahunje
27                     Sawai Mansingh Stadium, Jaipur
28                   M Chinnaswamy Stadium, Bangalore
29   Maharashtra Cricket Association Stadium, Gahunje
30                   M Chinnaswamy Stadium, Bangalore
31                             Feroz Shah Kotla, Delhi
32                              Eden Gardens, Kolkata
```

```
33     Maharani Usharaje Trust Cricket Ground, Indore
34   Maharashtra Cricket Association Stadium, Gahunje
35  Rajiv Gandhi International Stadium, Uppal, Hyd...
36                            Wankhede Stadium, Mumbai
37     Maharani Usharaje Trust Cricket Ground, Indore
38  Rajiv Gandhi International Stadium, Uppal, Hyd...
39                     Sawai Mansingh Stadium, Jaipur
40                              Eden Gardens, Kolkata
41                            Feroz Shah Kotla, Delhi
42                     Sawai Mansingh Stadium, Jaipur
43     Maharani Usharaje Trust Cricket Ground, Indore
44                            Feroz Shah Kotla, Delhi
45   Maharashtra Cricket Association Stadium, Gahunje
46                            Wankhede Stadium, Mumbai
47     Maharani Usharaje Trust Cricket Ground, Indore
48                              Eden Gardens, Kolkata
49                            Wankhede Stadium, Mumbai
50                  M Chinnaswamy Stadium, Bangalore
51                            Feroz Shah Kotla, Delhi
52                     Sawai Mansingh Stadium, Jaipur
53  Rajiv Gandhi International Stadium, Uppal, Hyd...
54                            Feroz Shah Kotla, Delhi
55   Maharashtra Cricket Association Stadium, Gahunje
56                            Wankhede Stadium, Mumbai
57                              Eden Gardens, Kolkata
58                              Eden Gardens, Kolkata
59                            Wankhede Stadium, Mumbai

                     home_team                    away_team  \
0                Mumbai Indians          Chennai Super Kings
1               Kings XI Punjab             Delhi Daredevils
2         Kolkata Knight Riders  Royal Challengers Bangalore
3                     Sunrisers             Rajasthan Royals
4           Chennai Super Kings        Kolkata Knight Riders
5              Rajasthan Royals             Delhi Daredevils
6                     Sunrisers               Mumbai Indians
7   Royal Challengers Bangalore              Kings XI Punjab
8                Mumbai Indians             Delhi Daredevils
9         Kolkata Knight Riders                    Sunrisers
10  Royal Challengers Bangalore             Rajasthan Royals
11              Kings XI Punjab          Chennai Super Kings
12        Kolkata Knight Riders             Delhi Daredevils
13               Mumbai Indians  Royal Challengers Bangalore
14             Rajasthan Royals        Kolkata Knight Riders
15              Kings XI Punjab                    Sunrisers
16          Chennai Super Kings             Rajasthan Royals
17        Kolkata Knight Riders              Kings XI Punjab
18  Royal Challengers Bangalore             Delhi Daredevils
```

```
19                   Sunrisers          Chennai Super Kings
20             Rajasthan Royals               Mumbai Indians
21             Delhi Daredevils              Kings XI Punjab
22               Mumbai Indians                    Sunrisers
23  Royal Challengers Bangalore          Chennai Super Kings
24                   Sunrisers              Kings XI Punjab
25             Delhi Daredevils        Kolkata Knight Riders
26           Chennai Super Kings               Mumbai Indians
27             Rajasthan Royals                    Sunrisers
28  Royal Challengers Bangalore        Kolkata Knight Riders
29           Chennai Super Kings             Delhi Daredevils
30  Royal Challengers Bangalore               Mumbai Indians
31             Delhi Daredevils             Rajasthan Royals
32        Kolkata Knight Riders          Chennai Super Kings
33              Kings XI Punjab               Mumbai Indians
34           Chennai Super Kings  Royal Challengers Bangalore
35                   Sunrisers             Delhi Daredevils
36               Mumbai Indians        Kolkata Knight Riders
37              Kings XI Punjab             Rajasthan Royals
38                   Sunrisers  Royal Challengers Bangalore
39             Rajasthan Royals              Kings XI Punjab
40        Kolkata Knight Riders               Mumbai Indians
41             Delhi Daredevils                    Sunrisers
42             Rajasthan Royals          Chennai Super Kings
43              Kings XI Punjab        Kolkata Knight Riders
44             Delhi Daredevils  Royal Challengers Bangalore
45           Chennai Super Kings                    Sunrisers
46               Mumbai Indians             Rajasthan Royals
47              Kings XI Punjab  Royal Challengers Bangalore
48        Kolkata Knight Riders             Rajasthan Royals
49               Mumbai Indians              Kings XI Punjab
50  Royal Challengers Bangalore                    Sunrisers
51             Delhi Daredevils          Chennai Super Kings
52             Rajasthan Royals  Royal Challengers Bangalore
53                   Sunrisers        Kolkata Knight Riders
54             Delhi Daredevils               Mumbai Indians
55           Chennai Super Kings              Kings XI Punjab
56           Chennai Super Kings                    Sunrisers
57        Kolkata Knight Riders             Rajasthan Royals
58        Kolkata Knight Riders                    Sunrisers
59           Chennai Super Kings                    Sunrisers

                    toss_winner toss_decision                    inn1team  \
0            Chennai Super Kings             f                Mumbai Indians
1                Kings XI Punjab             f              Delhi Daredevils
2          Kolkata Knight Riders             f  Royal Challengers Bangalore
3                     Sunrisers             f              Rajasthan Royals
4            Chennai Super Kings             f         Kolkata Knight Riders
```

5	Delhi Daredevils	f	Rajasthan Royals
6	Sunrisers	f	Mumbai Indians
7	Royal Challengers Bangalore	f	Kings XI Punjab
8	Delhi Daredevils	f	Mumbai Indians
9	Sunrisers	f	Kolkata Knight Riders
10	Royal Challengers Bangalore	f	Rajasthan Royals
11	Chennai Super Kings	f	Kings XI Punjab
12	Delhi Daredevils	f	Kolkata Knight Riders
13	Royal Challengers Bangalore	f	Mumbai Indians
14	Kolkata Knight Riders	f	Rajasthan Royals
15	Kings XI Punjab	b	Kings XI Punjab
16	Rajasthan Royals	f	Chennai Super Kings
17	Kings XI Punjab	f	Kolkata Knight Riders
18	Royal Challengers Bangalore	f	Delhi Daredevils
19	Sunrisers	f	Chennai Super Kings
20	Mumbai Indians	b	Mumbai Indians
21	Delhi Daredevils	f	Kings XI Punjab
22	Mumbai Indians	f	Sunrisers
23	Chennai Super Kings	f	Royal Challengers Bangalore
24	Kings XI Punjab	f	Sunrisers
25	Kolkata Knight Riders	f	Delhi Daredevils
26	Mumbai Indians	f	Chennai Super Kings
27	Sunrisers	b	Sunrisers
28	Kolkata Knight Riders	f	Royal Challengers Bangalore
29	Delhi Daredevils	f	Chennai Super Kings
30	Mumbai Indians	f	Royal Challengers Bangalore
31	Rajasthan Royals	f	Delhi Daredevils
32	Kolkata Knight Riders	f	Chennai Super Kings
33	Mumbai Indians	f	Kings XI Punjab
34	Chennai Super Kings	f	Royal Challengers Bangalore
35	Delhi Daredevils	b	Delhi Daredevils
36	Kolkata Knight Riders	f	Mumbai Indians
37	Kings XI Punjab	f	Rajasthan Royals
38	Royal Challengers Bangalore	f	Sunrisers
39	Rajasthan Royals	b	Rajasthan Royals
40	Kolkata Knight Riders	f	Mumbai Indians
41	Delhi Daredevils	b	Delhi Daredevils
42	Chennai Super Kings	b	Chennai Super Kings
43	Kings XI Punjab	f	Kolkata Knight Riders
44	Royal Challengers Bangalore	f	Delhi Daredevils
45	Chennai Super Kings	f	Sunrisers
46	Rajasthan Royals	f	Mumbai Indians
47	Royal Challengers Bangalore	f	Kings XI Punjab
48	Kolkata Knight Riders	f	Rajasthan Royals
49	Kings XI Punjab	f	Mumbai Indians
50	Sunrisers	f	Royal Challengers Bangalore
51	Chennai Super Kings	f	Delhi Daredevils
52	Rajasthan Royals	b	Rajasthan Royals

```
53                   Sunrisers             b                   Sunrisers
54             Delhi Daredevils             b             Delhi Daredevils
55           Chennai Super Kings            f              Kings XI Punjab
56           Chennai Super Kings            f                   Sunrisers
57              Rajasthan Royals            f        Kolkata Knight Riders
58        Kolkata Knight Riders             f                   Sunrisers
59           Chennai Super Kings            f                   Sunrisers

    innings1  ...  adjusted_target_indicator  adjusted_target team1_overs  \
0        165  ...                         n                0        20.0
1        166  ...                         n                0        20.0
2        176  ...                         n                0        20.0
3        125  ...                         n                0        20.0
4        202  ...                         n                0        20.0
5        153  ...                         y               71        17.5
6        147  ...                         n                0        20.0
7        155  ...                         n                0        20.0
8        194  ...                         n                0        20.0
9        138  ...                         n                0        20.0
10       217  ...                         n                0        20.0
11       197  ...                         n                0        20.0
12       200  ...                         n                0        20.0
13       213  ...                         n                0        20.0
14       160  ...                         n                0        20.0
15       193  ...                         n                0        20.0
16       204  ...                         n                0        20.0
17       191  ...                         y              125        20.0
18       174  ...                         n                0        20.0
19       182  ...                         n                0        20.0
20       167  ...                         n                0        20.0
21       143  ...                         n                0        20.0
22       118  ...                         n                0        20.0
23       205  ...                         n                0        20.0
24       132  ...                         n                0        20.0
25       219  ...                         n                0        20.0
26       169  ...                         n                0        20.0
27       151  ...                         n                0        20.0
28       175  ...                         n                0        20.0
29       211  ...                         n                0        20.0
30       167  ...                         n                0        20.0
31       196  ...                         y              151        17.1
32       177  ...                         n                0        20.0
33       174  ...                         n                0        20.0
34       127  ...                         n                0        20.0
35       163  ...                         n                0        20.0
36       181  ...                         n                0        20.0
37       152  ...                         n                0        20.0
38       146  ...                         n                0        20.0
```

39	158	...	n	0	20.0
40	210	...	n	0	20.0
41	187	...	n	0	20.0
42	176	...	n	0	20.0
43	245	...	n	0	20.0
44	181	...	n	0	20.0
45	179	...	n	0	20.0
46	168	...	n	0	20.0
47	88	...	n	0	20.0
48	142	...	n	0	20.0
49	186	...	n	0	20.0
50	218	...	n	0	20.0
51	162	...	n	0	20.0
52	164	...	n	0	20.0
53	172	...	n	0	20.0
54	174	...	n	0	20.0
55	153	...	n	0	20.0
56	139	...	n	0	20.0
57	169	...	n	0	20.0
58	174	...	n	0	20.0
59	178	...	n	0	20.0

	team2_overs	mom_player_id	mom_player	scoring_status	result_type	\
0	20	44613	DJ Bravo	live bbb	ww	
1	20	170187	KL Rahul	live bbb	ww	
2	20	412485	N Rana	live bbb	ww	
3	20	15627	S Dhawan	live bbb	ww	
4	20	119895	SW Billings	live bbb	ww	
5	6	218180	SV Samson	live bbb	wr	
6	20	1738090	Rashid Khan	live bbb	ww	
7	20	356989	UT Yadav	live bbb	ww	
8	20	157435	JJ Roy	live bbb	ww	
9	20	1073273	B Stanlake	live bbb	ww	
10	20	218180	SV Samson	live bbb	wr	
11	20	7537	CH Gayle	live bbb	wr	
12	20	412485	N Rana	live bbb	wr	
13	20	74266	RG Sharma	live bbb	wr	
14	20	412485	N Rana	live bbb	ww	
15	20	7537	CH Gayle	live bbb	wr	
16	20	10016	SR Watson	live bbb	wr	
17	13	170187	KL Rahul	live bbb	ww	
18	20	45848	AB de Villiers	live bbb	ww	
19	20	8931	AT Rayudu	live bbb	wr	
20	20	1422835	JC Archer	live bbb	ww	
21	20	1047395	A Rajpoot	live bbb	wr	
22	20	1738090	Rashid Khan	live bbb	wr	
23	20	7561	MS Dhoni	live bbb	ww	
24	20	1047395	A Rajpoot	live bbb	wr	

25	20	967227	SS Iyer	live bbb	wr
26	20	74266	RG Sharma	live bbb	ww
27	20	159052	KS Williamson	live bbb	wr
28	20	158662	CA Lynn	live bbb	ww
29	20	10016	SR Watson	live bbb	wr
30	20	97576	TG Southee	live bbb	wr
31	12	1065017	RR Pant	live bbb	wr
32	20	97646	SP Narine	live bbb	ww
33	20	264193	SA Yadav	live bbb	ww
34	20	94280	RA Jadeja	live bbb	ww
35	20	1738090	Rashid Khan	live bbb	ww
36	20	392753	HH Pandya	live bbb	wr
37	20	2128581	Mujeeb Zadran	live bbb	ww
38	20	159052	KS Williamson	live bbb	wr
39	20	136065	JC Buttler	live bbb	wr
40	20	1377783	IP Kishan	live bbb	wr
41	20	15627	S Dhawan	live bbb	ww
42	20	136065	JC Buttler	live bbb	ww
43	20	97646	SP Narine	live bbb	wr
44	20	45848	AB de Villiers	live bbb	ww
45	20	8931	AT Rayudu	live bbb	ww
46	20	136065	JC Buttler	live bbb	ww
47	20	356989	UT Yadav	live bbb	ww
48	20	404424	K Yadav	live bbb	ww
49	20	1196767	JJ Bumrah	live bbb	wr
50	20	45848	AB de Villiers	live bbb	wr
51	20	250417	HV Patel	live bbb	wr
52	20	171459	R Shreyas Gopal	live bbb	wr
53	20	158662	CA Lynn	live bbb	ww
54	20	12353	A Mishra	live bbb	wr
55	20	1214926	LT Ngidi	live bbb	ww
56	20	60234	F du Plessis	live bbb	ww
57	20	158041	AD Russell	live bbb	wr
58	20	1738090	Rashid Khan	live bbb	wr
59	20	10016	SR Watson	live bbb	ww

	result_margin	winning_team
0	1	Chennai Super Kings
1	6	Kings XI Punjab
2	4	Kolkata Knight Riders
3	9	Sunrisers
4	5	Chennai Super Kings
5	10	Rajasthan Royals
6	1	Sunrisers
7	4	Royal Challengers Bangalore
8	7	Delhi Daredevils
9	5	Sunrisers
10	19	Rajasthan Royals

11	4	Kings XI Punjab
12	71	Kolkata Knight Riders
13	46	Mumbai Indians
14	7	Kolkata Knight Riders
15	15	Kings XI Punjab
16	64	Chennai Super Kings
17	9	Kings XI Punjab
18	6	Royal Challengers Bangalore
19	4	Chennai Super Kings
20	3	Rajasthan Royals
21	4	Kings XI Punjab
22	31	Sunrisers
23	5	Chennai Super Kings
24	13	Sunrisers
25	55	Delhi Daredevils
26	8	Mumbai Indians
27	11	Sunrisers
28	6	Kolkata Knight Riders
29	13	Chennai Super Kings
30	14	Royal Challengers Bangalore
31	4	Delhi Daredevils
32	6	Kolkata Knight Riders
33	6	Mumbai Indians
34	6	Chennai Super Kings
35	7	Sunrisers
36	13	Mumbai Indians
37	6	Kings XI Punjab
38	5	Sunrisers
39	15	Rajasthan Royals
40	102	Mumbai Indians
41	9	Sunrisers
42	4	Rajasthan Royals
43	31	Kolkata Knight Riders
44	5	Royal Challengers Bangalore
45	8	Chennai Super Kings
46	7	Rajasthan Royals
47	10	Royal Challengers Bangalore
48	6	Kolkata Knight Riders
49	3	Mumbai Indians
50	14	Royal Challengers Bangalore
51	34	Delhi Daredevils
52	30	Rajasthan Royals
53	5	Kolkata Knight Riders
54	11	Delhi Daredevils
55	5	Chennai Super Kings
56	2	Chennai Super Kings
57	25	Kolkata Knight Riders
58	14	Sunrisers

```
59              8          Chennai Super Kings

[60 rows x 27 columns]
```

In [22]:
```
# This cell compelete a number tasks. First we identify when the home team is the wins
# is the winner. Next we identify the runs scored by the home team and the away team
# nine innings for each team, in T20 cricket each team gets only one inning, and once
# team has its inning). Finally, we include a counter which we can add up to give tot

IPL18["hWin"] = np.where(IPL18['home_team']== IPL18['winning_team'],1,0)
IPL18["aWin"] = np.where(IPL18['away_team']== IPL18['winning_team'],1,0)
IPL18["hRuns"] = np.where(IPL18['home_team']== IPL18['inn1team'], IPL18['innings1'], I
IPL18["aRuns"] = np.where(IPL18['away_team']== IPL18['inn1team'], IPL18['innings1'], I
IPL18["count"] = 1
```

In [23]:
```
# Now we use a .groupby command to aggregate the performance of home teams during the
# to see how similar the commands are.

IPLHome = IPL18.groupby("home_team")['count','hWin','hRuns', 'aRuns'].sum().reset_inde
IPLHome = IPLHome.rename(columns = {'home_team':'team','count':'Ph', 'hWin':'hWinH',
IPLHome
```

Out[23]:

	team	Ph	hWinH	hRuns	aRuns
0	Chennai Super Kings	9	8	1577	1486
1	Delhi Daredevils	7	4	1258	1122
2	Kings XI Punjab	7	4	1188	1202
3	Kolkata Knight Riders	9	5	1468	1417
4	Mumbai Indians	7	3	1194	1171
5	Rajasthan Royals	7	5	1120	994
6	Royal Challengers Bangalore	7	4	1298	1286
7	Sunrisers	7	5	1070	1050

In [25]:
```
# Now we aggregate the performance of away teams in a different df.

IPLAway = IPL18.groupby("away_team")['count','aWin','hRuns', 'aRuns'].sum().reset_inde
IPLAway = IPLAway.rename(columns = {'away_team':'team','count':'Pa', 'hWin':'hWinH',
IPLAway
```

Out[25]:

	team	Pa	aWinH	hRuns	aRuns
0	Chennai Super Kings	7	3	1264	1232
1	Delhi Daredevils	7	1	1265	1085
2	Kings XI Punjab	7	2	1124	1022
3	Kolkata Knight Riders	7	4	1326	1291
4	Mumbai Indians	7	3	1111	1186
5	Rajasthan Royals	8	2	1362	1237
6	Royal Challengers Bangalore	7	2	1097	1024
7	Sunrisers	10	5	1624	1651

In [26]:
```
# how we merge the two dfs to obtain a full record for each team across the season.
```

```
IPL18 = IPLHome.merge(IPLAway, on = "team")
IPL18
```

Out[26]:

	team	Ph	hWinH	hRuns_x	aRuns_x	Pa	aWinH	hRuns_y	aRuns_y
0	Chennai Super Kings	9	8	1577	1486	7	3	1264	1232
1	Delhi Daredevils	7	4	1258	1122	7	1	1265	1085
2	Kings XI Punjab	7	4	1188	1202	7	2	1124	1022
3	Kolkata Knight Riders	9	5	1468	1417	7	4	1326	1291
4	Mumbai Indians	7	3	1194	1171	7	3	1111	1186
5	Rajasthan Royals	7	5	1120	994	8	2	1362	1237
6	Royal Challengers Bangalore	7	4	1298	1286	7	2	1097	1024
7	Sunrisers	7	5	1070	1050	10	5	1624	1651

In [27]:
```
# We now aggregate the home and away data for wins, games played and runs

IPL18['Wins'] = IPL18['hWinH']+IPL18['aWinH']
IPL18['GP'] = IPL18['Ph']+IPL18['Pa']
IPL18['Truns'] = IPL18['hRuns_x']+IPL18['aRuns_y']
IPL18['RAgainst'] = IPL18['hRuns_y']+IPL18['aRuns_x']

IPL18
```

Out[27]:

	team	Ph	hWinH	hRuns_x	aRuns_x	Pa	aWinH	hRuns_y	aRuns_y	Wins	GP	Truns	RAgainst
0	Chennai Super Kings	9	8	1577	1486	7	3	1264	1232	11	16	2809	2750
1	Delhi Daredevils	7	4	1258	1122	7	1	1265	1085	5	14	2343	2387
2	Kings XI Punjab	7	4	1188	1202	7	2	1124	1022	6	14	2210	2326
3	Kolkata Knight Riders	9	5	1468	1417	7	4	1326	1291	9	16	2759	2743
4	Mumbai Indians	7	3	1194	1171	7	3	1111	1186	6	14	2380	2282
5	Rajasthan Royals	7	5	1120	994	8	2	1362	1237	7	15	2482	2356
6	Royal Challengers Bangalore	7	4	1298	1286	7	2	1097	1024	6	14	2395	2310
7	Sunrisers	7	5	1070	1050	10	5	1624	1651	10	17	2694	2701

```
5     1362     1237     7  15   2357      2356
6     1097     1024     6  14   2322      2383
7     1624     1651    10  17   2721      2674
```

In [29]:
```
# The last step in organizing the data is to create variables for win percentage (wpc)

IPL18['WPC'] = IPL18['Wins']/IPL18['GP']
IPL18['PEX'] = (IPL18['Truns']**2)/(IPL18['Truns']**2+IPL18['RAgainst']**2)
IPL18
```

Out[29]:

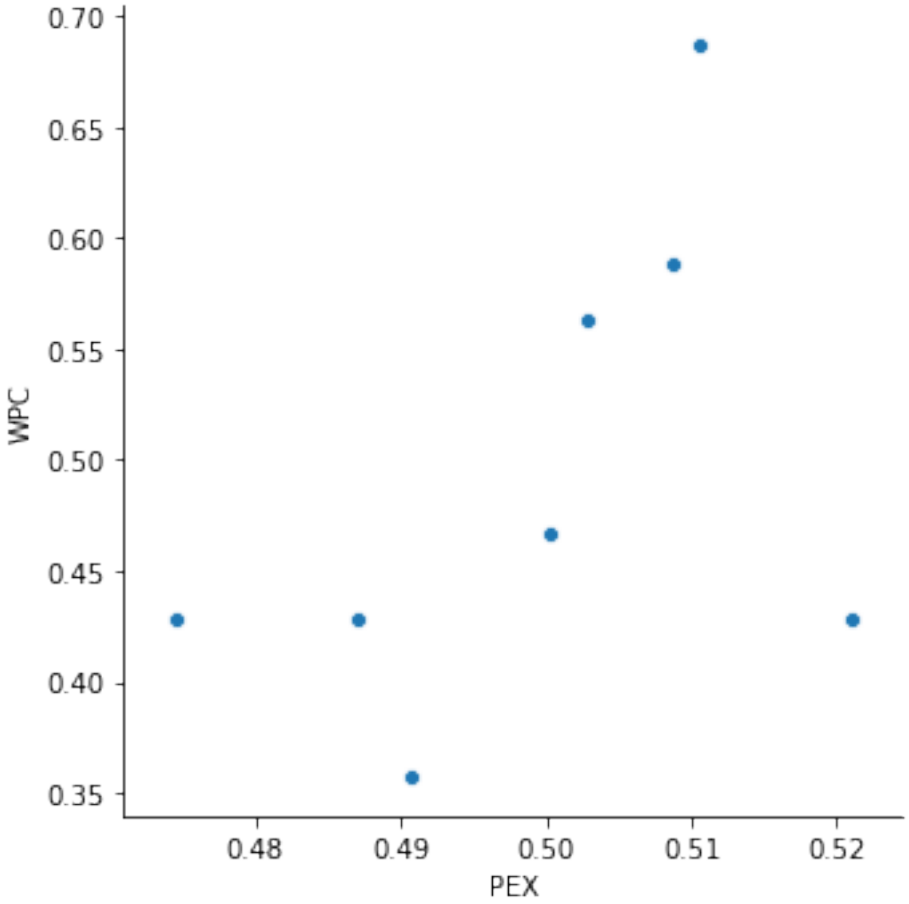
	team	Ph	hWinH	hRuns_x	aRuns_x	Pa	aWinH	\
0	Chennai Super Kings	9	8	1577	1486	7	3	
1	Delhi Daredevils	7	4	1258	1122	7	1	
2	Kings XI Punjab	7	4	1188	1202	7	2	
3	Kolkata Knight Riders	9	5	1468	1417	7	4	
4	Mumbai Indians	7	3	1194	1171	7	3	
5	Rajasthan Royals	7	5	1120	994	8	2	
6	Royal Challengers Bangalore	7	4	1298	1286	7	2	
7	Sunrisers	7	5	1070	1050	10	5	

	hRuns_y	aRuns_y	Wins	GP	Truns	RAgainst	WPC	PEX
0	1264	1232	11	16	2809	2750	0.687500	0.510612
1	1265	1085	5	14	2343	2387	0.357143	0.490698
2	1124	1022	6	14	2210	2326	0.428571	0.474444
3	1326	1291	9	16	2759	2743	0.562500	0.502908
4	1111	1186	6	14	2380	2282	0.428571	0.521012
5	1362	1237	7	15	2357	2356	0.466667	0.500212
6	1097	1024	6	14	2322	2383	0.428571	0.487037
7	1624	1651	10	17	2721	2674	0.588235	0.508711

In [31]:
```
# Having prepared the data, we are now ready to examine it. First, we generate and xy
# Unlike the MLB case, we can see that there is a very weak correlation between win p

sns.relplot(x = 'PEX', y = 'WPC', data = IPL18)
```

Out[31]:
```
<seaborn.axisgrid.FacetGrid at 0x7f0a2317c160>
```



1.1 Self test

run sns.relplot again, but this time write y="W" instead of y="wpc". What do you find? Does it make a difference?

2 Running a regression

We now run the same regression as we did for the MLB data:

wpc = Intercept + coef x pyth

This time, while coefficient on pyth is positive - implying that a higher Pythagorean Expectation leads to a large win percentage, the standard error is also very large, and the t statistic of 1.353 implies a p-value of 0.225- well above the usual threshold of 0.050, which means that the coefficient estimate is in fact insignificantly different from zero.

```
In [32]: pex_lm = smf.ols(formula = 'WPC~PEX', data = IPL18).fit()
         pex_lm.summary()
```

```
/opt/conda/lib/python3.6/site-packages/scipy/stats/stats.py:1416: UserWarning: kurtosistest onl
  "anyway, n=%i" % int(n))
```

Out[32]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    WPC   R-squared:                       0.234
Model:                            OLS   Adj. R-squared:                  0.106
Method:                 Least Squares   F-statistic:                     1.830
Date:                Tue, 16 Aug 2022   Prob (F-statistic):              0.225
Time:                        17:00:47   Log-Likelihood:                 7.9710
No. Observations:                   8   AIC:                            -11.94
Df Residuals:                       6   BIC:                            -11.78
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -1.2807      1.312     -0.976      0.367      -4.491       1.929
PEX            3.5522      2.626      1.353      0.225      -2.872       9.977
==============================================================================
Omnibus:                        0.002   Durbin-Watson:                   2.254
Prob(Omnibus):                  0.999   Jarque-Bera (JB):                0.217
Skew:                           0.014   Prob(JB):                        0.897
Kurtosis:                       2.193   Cond. No.                         89.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly speci
"""
```

2.1 Self test

Run the regression above but instead write 'wpc ~ W' instead of 'wpc ~ pyth' in the line starting pyth_lm. What difference does this make?

3 Conclusion

Why did the Pythagorean model produce a good fit for the baseball data but not for the cricket data? An obvious explanation is that there is some difference between the two sports which makes the model appropriate for one but not the other. For example, in cricket, the team batting second need only score one more run than the opponent to win, and so the inning ends if it reaches this milestone. If the team batting second is the winning team, then the gap in the scores will be small. However, if the team batting first can get all ten wickets cheaply, then the gap in scores could be very large. In our data the average runs difference when the team batting second won was 2, and

when the team batting first won was 30. This might explain why the Pythagorean Expectation is not a good guide to winning in the IPL.

But there could be more basic statistical explanations. For MLB we had averages for 30 teams, each of which played about 160 games. Random variations are likely to be smoothed out when analyzing data on this scale. For the IPL we had only 8 teams, most of whom played only 14 games - so there is a much greater chance that random variations could have overwhelmed the Pythagorean model if it were correct.

Anyone interested in pursuing this further, might try two things. First, analyze games where the winning team bats first or second separately. Second, find data covering more seasons (not difficult to find online) in order to generate a much larger sample.

For now, however, we are going to move on and look at another sport: basketball.

In [33]:
```
pex_lm = smf.ols(formula = 'WPC~Wins', data = IPL18).fit()
pex_lm.summary()
```

```
/opt/conda/lib/python3.6/site-packages/scipy/stats/stats.py:1416: UserWarning: kurtosistest onl
  "anyway, n=%i" % int(n))
```

Out[33]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    WPC   R-squared:                       0.979
Model:                            OLS   Adj. R-squared:                  0.976
Method:                 Least Squares   F-statistic:                     285.5
Date:                Tue, 16 Aug 2022   Prob (F-statistic):           2.75e-06
Time:                        17:02:44   Log-Likelihood:                 22.439
No. Observations:                   8   AIC:                            -40.88
Df Residuals:                       6   BIC:                            -40.72
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.1260      0.023      5.587      0.001       0.071       0.181
Wins           0.0490      0.003     16.897      0.000       0.042       0.056
==============================================================================
Omnibus:                        0.632   Durbin-Watson:                   2.172
Prob(Omnibus):                  0.729   Jarque-Bera (JB):                0.341
Skew:                          -0.426   Prob(JB):                        0.843
Kurtosis:                       2.454   Cond. No.                         29.8
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly speci
"""
```

In [ ]: