2018-2022 MLB Umpire Performance Analysis

Data Acquisition:

In this project, we are looking at the pitch call-strike accuracy for the MLB umpires during 2018-2022 season. We download the entire 5-year pitch information from pybaseball API and cut down to only called balls and called strikes. Then we apply two functions to check if umpires made the correct call during the game. First function checks if the ball landed in the home plate, and the second function runs 1000 simulations to test if umpire can make the call correctly 95% of the time. (data processing.py)

Strikes vs Balls:

Our data consists of over 1.6 million total pitches, 1.5 million of which are called correctly. This gives us a 5-year Bad Call Ratio of 7.6%.

```
len(df['pitch_type'])
```
```
1624283
```

```
display(df["call"].value_counts())
```
```
correct      1500923
incorrect     123360
Name: call, dtype: int64
```

```
print('5 year BCR: ',df['call'].mean())
```
```
5 year BCR:  0.07594735646435996
```

We then break down the data into balls and strikes. Roughly two thirds of the total pitches are balls and rest strikes. We see that strikes have a BCR of 13.6% and balls have a BCR of 4.6%.

```
df['description'].value_counts()
```
```
ball            1087477
called_strike    536806
Name: description, dtype: int64
```

strike bad call ratio	ball bad call ratio
0.13563372987634267	0.04648466128479039

This means that of if an umpire calls a pitch a strike, 13.56% of the time that this pitch is indeed a ball, which implies that umpire favors pitchers more than batters.

Pitch Types:

During the 5 years, there are total of 16 different pitch types and their corresponding pitch numbers and bad call ratios are as follows.

```
pt.groupby("pitch_type").mean()
```

	call
pitch_type	
CH	0.057672
CS	0.042345
CU	0.067717
EP	0.110070
FA	0.077330
FC	0.080290
FF	0.082587
FO	0.066667
FS	0.050976
KC	0.066317
KN	0.095764
SC	0.227273
SI	0.087315
SL	0.066903
ST	0.064982
SV	0.067629

```
df['pitch_type'].value_counts()
```

```
FF    587488
SI    284039
SL    263410
CH    158516
CU    140128
FC    101956
KC     40095
FS     20794
ST     20067
SV      4850
FA      1513
KN       543
EP       427
CS       307
FO        45
SC        22
```

After removing 5 outliers with the less than 1000 occurrence, we see that umpires have the highest bad call ratio at FF (0.082587) and SI (0.087315), while having the lowest bad call ratio at CH (0.057672) and FS (0.050976). The average bad call ratio for the 11 pitch types is 7%, and four-seam fastballs (FF) are 18% more likely than average to be called incorrectly and sinkers (SI) are 24.8% more likely to be called incorrectly.
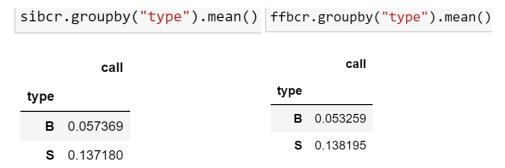
Pitch Zones:

The home plate is divided into 9 equal parts, with zone 11, 12, 13, 14 outside the Homeplate on the four corners. Each pitch lands in one of the 13 zones and their corresponding occurrences and Bad Call Ratios are as follows.

```
zonebcr.groupby("zone").mean()
```

	call
zone	
1.0	0.203636
2.0	0.097626
3.0	0.208971
4.0	0.029866
5.0	0.000655
6.0	0.036987
7.0	0.114895
8.0	0.077530
9.0	0.152035
11.0	0.055528
12.0	0.075381
13.0	0.078378
14.0	0.058418

```
df['zone'].value_counts()

14.0    391404
13.0    265341
11.0    260498
12.0    194770
 9.0     75555
 8.0     66955
 7.0     62657
 6.0     62589
 4.0     58596
 5.0     54990
 1.0     47305
 3.0     42896
 2.0     40727
```

The thirteen strikes zones have a average Bad Call Ratio of 9.2%. Zones 1 and 3 are 2.22 times and 2.28 times more likely than average to be called incorrectly respectively. Zones 7 and 9 are 26% and 66% more likely than average to be called incorrectly. It appears that umpires are better at calling pitches that lands at the bottom of the home plate than the ones that lands at the top.

Sinkers and Fast Balls:

As previously stated, Sinkers and Fast Balls have the highest Bad Call Ratios amongst all the pitch types. Their respective Ball and Strike Bad Call Ratios are as follows.

```
sibcr.groupby("type").mean()
```     ```
ffbcr.groupby("type").mean()
```

|      | call |
|------|------|
| type |      |
| B    | 0.057369 |
| S    | 0.137180 |

|      | call |
|------|------|
| type |      |
| B    | 0.053259 |
| S    | 0.138195 |

If a pitcher throws a Sinker, 5.7% of the chances that the called ball will be a strike, and 13.7% of the chances that the called strikes should be a ball. If a pitcher throws a Fast Ball, 5.3% of the chances that the called ball will be a strike, and 13.8% of the chances that the called strikes should be a ball.