The Astounding Salemi Stats

The report examines three new baseball stats that will revolutionize baseball analysis. There are three Salemi Stats:

BasesPerHit (BPH)

We divide the total bases by the number of hits to get the average number of bases we get from every hit.

$$\frac{(H\ +\ 2B\ +\ 2\ *\ 3B\ +\ 3\ *\ \mathrm{HR}\ )}{H}$$

Productive Strikouts (PSO)

Calculate the hits, walks, and hit by pitches per strikeout.

$$\frac{(H\ +\ \mathrm{BB}\ +\ \mathrm{HBP}\ )}{\mathrm{SO}}$$

Speed Trial (ST)

We calculate the stolen bases - 2 * caught stealing per H

$$\frac{(\mathrm{SB}\ -\ 2\ *\ \mathrm{CS}\ )}{H}$$

Read the data

```
Warning: Column headers from the file were modified to make them valid MATLAB identifiers
before creating variable names for the table. The original column headers are saved in the
VariableDescriptions property.
Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.
```

Remove the team names

Create the statistics

BPH

Code hidden

PSO

Code hidden

ST

Code hidden

ans = 5x4 table

	R	BPH	PSO	ST
1	807	1.6533	1.4277	0.0472
2	772	1.6423	1.4534	0.0418
3	641	1.6139	1.2198	0.0321

	R	BPH	PSO	ST
4	662	1.5250	1.6681	0.0573
5	661	1.6176	1.2553	0.0302

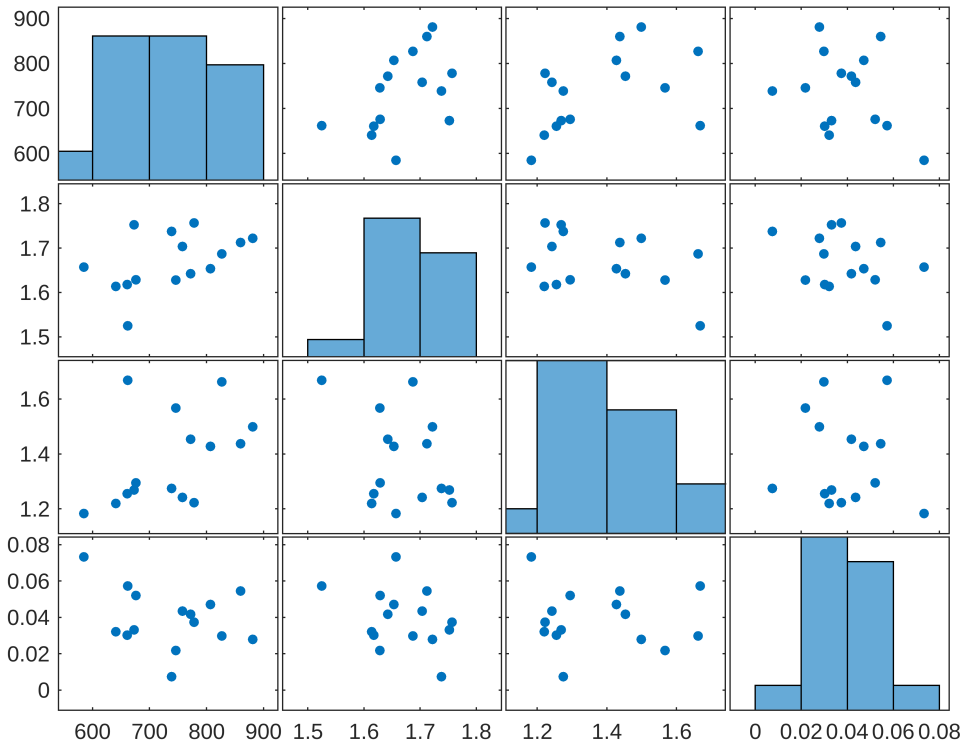
Find the best statistic

We'll use the plot matrix to see the greatness of the new stats and then calculate the coefficients.

Plot Matrix

First we look at a plot matrix using the four columns in the matrix above.

Code hidden



Hmm . . . this is not looking so good.

Correlation Coefficients

Now calculate the Correlation Coeffiicients with the order R BPH PSO ST.

Code hidden

```
ans = 4×4
    1.0000    0.4646    0.4703   -0.2927
    0.4646    1.0000   -0.3567   -0.3445
    0.4703   -0.3567    1.0000   -0.0657
   -0.2927   -0.3445   -0.0657    1.0000
```

Conclusion

My new statistics are all terrible! Far worse than batting average.

The best of the worst was Productive Strike Outs (PSO)

$$\frac{(H + \text{BB} + \text{HBP})}{\text{SO}}$$

This statistic had a .4703 correlation to Runs Scored (R).