Data 8 Connector: Sports Analytics

...

Review - 02/13/18

Course stuff

- Project proposals: due end of day today. Email to

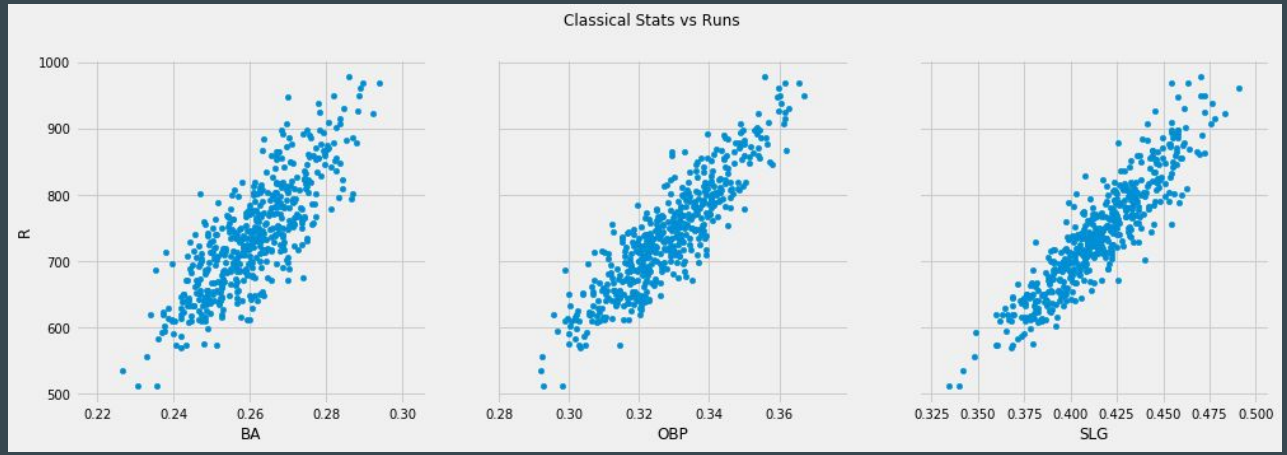
  apapanicolaou@berkeley.edu

- Homework: 4th Down Bot still coming
- Still evolving: to avoid not enough of other sports, will take plenty of time to introduce related concepts from other sports/fields
  - Not quite to same depth due to limitations (you'll see why)
  - Last week: shooting %s in basketball
  - Today will actually be financial markets
  - In the future: Basketball expected value model, DVOA for football, etc

Recap

Classic Stats, Run Scoring, and Run Expectancy

The classic stats had a good correlation with run scoring



In some fields, that level of correlation would be unheard of. Of course (finance sometimes), there's a higher standard in baseball

Classic Stats, Run Scoring, and Run Expectancy

OBP and SLG served as better predictors of run scoring than BA

- Higher correlation, smaller errors

Just one thing: try not to be a snob.

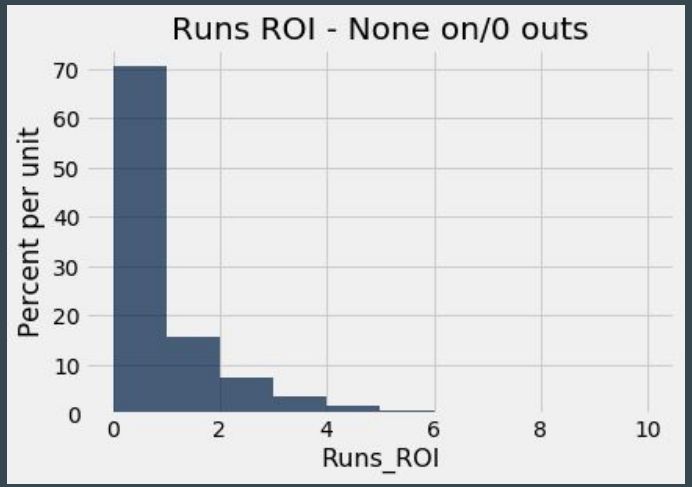
- Without anything else, the classic stats are okay
- They're a good place to start but if you can dig deeper, do so.

Still our goal: find better weight values for events

Classic Stats, Run Scoring, and Run Expectancy

We introduced run expectancy

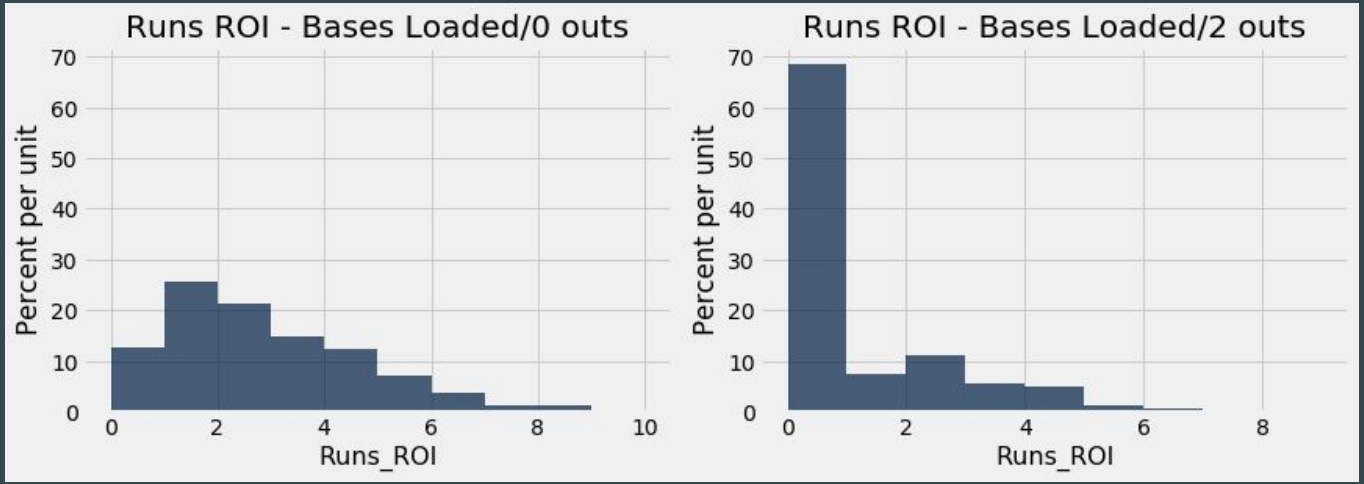
➔ Given a number of outs and a base state, how many runs do you expect to score?



Classic Stats, Run Scoring, and Run Expectancy

We introduced run expectancy

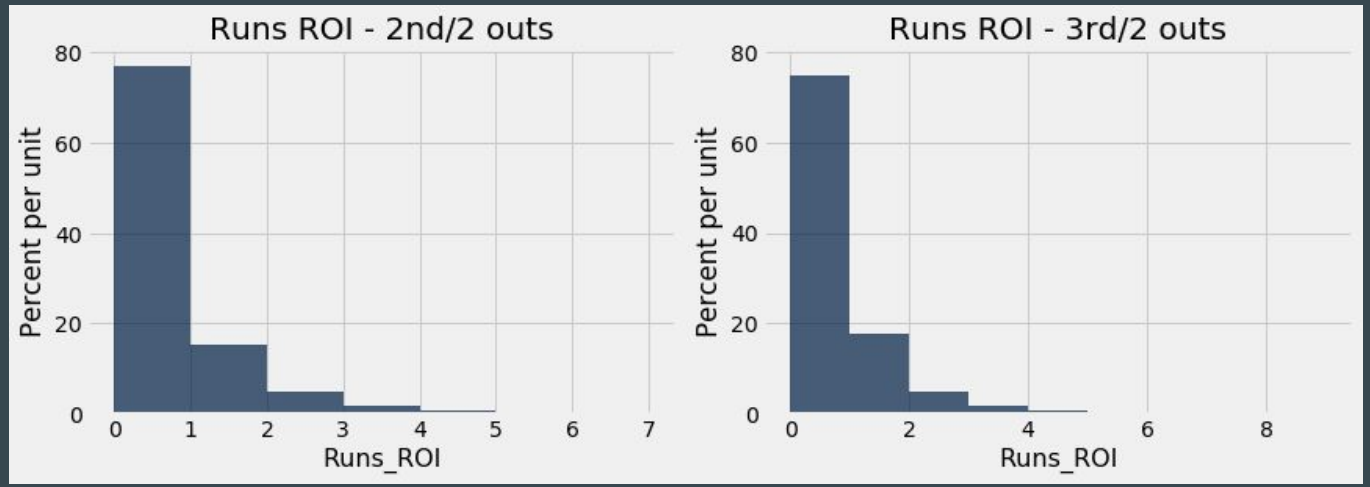
➔ Given a number of outs and a base state, how many runs do you expect to score?



Classic Stats, Run Scoring, and Run Expectancy

We introduced run expectancy

➔ Given a number of outs and a base state, how many runs do you expect to score?



Classic Stats, Run Scoring, and Run Expectancy

We introduced run expectancy

- ➔ Average over all values and you get run expectancy
- ➔ It's an expected value under completely average circumstances: average hitters, average pitcher, etc

Classic Stats, Run Scoring, and Run Expectancy

In the homework: win probability or win expectancy

- ➔ Compute probability of scoring at least one run: chances of winning in a late tie game
- ➔ More generally, you can look at score differences and other innings

Strategies

- ➔ Stolen bases and bunting: can use RE or WP to figure out what to do
- ➔ WP more general for choosing late game strategies than RE

Inning	Top/Bottom	Score	Outs	1B	2B	3B	WE
7	Bottom	-1	0				0.353
7	Bottom	-1	0	1st			0.431
7	Bottom	-1	0	1st	2nd		0.545
7	Bottom	-1	0	1st	2nd	3rd	0.687
7	Bottom	-1	0	1st		3rd	0.612
7	Bottom	-1	0		2nd		0.487
7	Bottom	-1	0		2nd	3rd	0.656
7	Bottom	-1	0			3rd	0.545
Inning	Top/Bottom	Score	Outs	1B	2B	3B	WE
7	Bottom	-1	1				0.305
7	Bottom	-1	1	1st			0.360
7	Bottom	-1	1	1st	2nd		0.438
7	Bottom	-1	1	1st	2nd	3rd	0.561
7	Bottom	-1	1	1st		3rd	0.504
7	Bottom	-1	1		2nd		0.399
7	Bottom	-1	1		2nd	3rd	0.546
7	Bottom	-1	1			3rd	0.465
Inning	Top/Bottom	Score	Outs	1B	2B	3B	WE
7	Bottom	-1	2				0.268
7	Bottom	-1	2	1st			0.296
7	Bottom	-1	2	1st	2nd		0.341
7	Bottom	-1	2	1st	2nd	3rd	0.406
7	Bottom	-1	2	1st		3rd	0.359
7	Bottom	-1	2		2nd		0.322
7	Bottom	-1	2		2nd	3rd	0.382
7	Bottom	-1	2			3rd	0.336

Related Concept: Financial Markets & Algorithmic Trading

Expected value modeling is pervasive and one area where it can help in a volatile system is financial markets

Related Concept: Financial Markets & Algorithmic Trading

How buying and selling of stocks/futures/currencies works
(a bit stylized but generally true)

- ➔ You want to buy $50k Apple
- ➔ You have access to the venue/exchange where buying and selling happens (in reality, you don't but let's pretend you're a hedge fund)
- ➔ The "market" is a bunch of banks/hedge funds etc in the venue offering up prices to buy and sell to you or anyone else
- ➔ Your order matches with what's there in the venue: if you aren't willing to pay the quoted prices, there is no match.
- ➔ All the orders coming from various places cause the price to fluctuate

Related Concept: Financial Markets & Algorithmic Trading

It's a pretty simple market and not that much different from what we're used to

- ➔ Or even what's depicted in Trading Places

But here's where it gets complicated

- ➔ Speed plays a factor and sometimes you need to place very large orders
- ➔ Speed (this is not a judgment!): the price may move before you get there
  You need to model the probability/expectation of actually filling the order
- ➔ Large orders (very large, much larger than what's quoted): they need to be split over minutes/hours/days/weeks
  Buying makes the price go up, selling makes the price go down
  You need to model the expected effects of your order

Related Concept: Financial Markets & Algorithmic Trading

Expected value modeling

- ➔ Can help you with what to expect when placing the order
- ➔ It's not perfect because there is so much happening that causes the price to fluctuate

Related Concept: Financial Markets & Algorithmic Trading

A stylized depiction

- ➔ Blue shows the modeled price movement based on every other participant
- ➔ Red shows the order's pressure pushing the price up
- ➔ Yellow is the actual price including your effect

Related Concept: Financial Markets & Algorithmic Trading

Price you see

- Information that leads into a model that then creates a signal to trade

Price you expect

- Your model for the actual trading leads to an expectation of the price (which can affect your willingness to trade)

Price you get

- Mismatch between model and observed
- Can tell you if your execution is bad (maybe you go through a broker or your IT built a bad/slow system)
- Or your model is weak.