Applying Empirical Bayesian Techniques to Illinois Traffic Stop Data

--Bryan Berend (he/him/his)

--Senior Applied Data Scientist @ Civis Analytics

--satRday Chicago, April 27, 2019

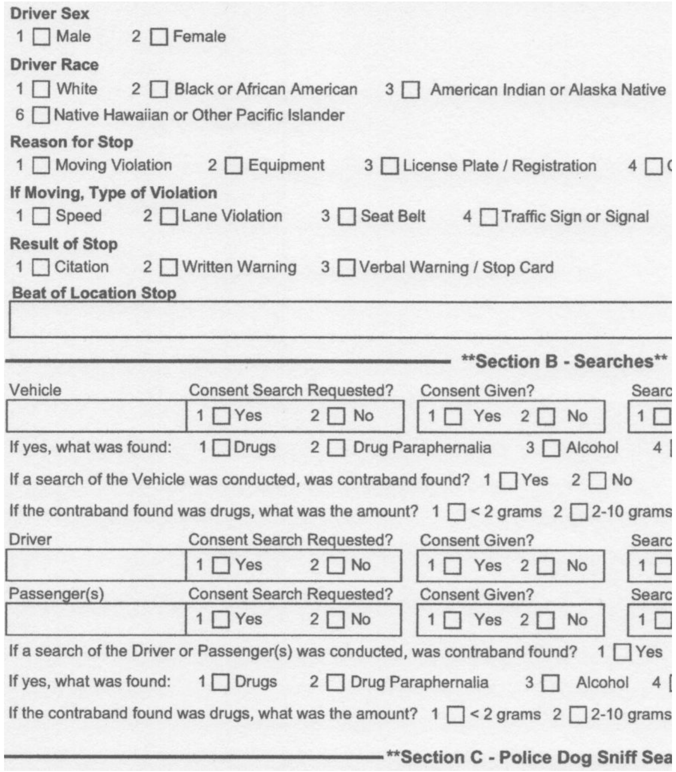
Agenda

- Overview of IL traffic stop data
- Introduction to Empirical Bayesian techniques
- Use EB techniques on IL traffic stop data
- Takeaways & caveats
- Q&A

Illinois Traffic Stop Data - Background

- Since 2004, the Illinois Traffic and Pedestrian Stop Statistical Study Act has required Illinois law enforcement to document and report traffic stops
- Data collected includes the driver’s info, officer’s info, what happened during the stop, if any dog sniffs or searches occured, and the result of the stop
- Data visualizer Mollie Pettit and and data scientist Chris Kucharczyk cleaned and analyzed this data; their interactive findings can be found at https://illinoistrafficstops.com/

A partial screenshot of the form officers must fill out for each traffic stop



Driver Sex
1 ☐ Male 2 ☐ Female
Driver Race
1 ☐ White 2 ☐ Black or African American 3 ☐ American Indian or Alaska Native
6 ☐ Native Hawaiian or Other Pacific Islander
Reason for Stop
1 ☐ Moving Violation 2 ☐ Equipment 3 ☐ License Plate / Registration 4 ☐
If Moving, Type of Violation
1 ☐ Speed 2 ☐ Lane Violation 3 ☐ Seat Belt 4 ☐ Traffic Sign or Signal
Result of Stop
1 ☐ Citation 2 ☐ Written Warning 3 ☐ Verbal Warning / Stop Card
Beat of Location Stop

Section B - Searches

Vehicle | Consent Search Requested? 1 ☐ Yes 2 ☐ No | Consent Given? 1 ☐ Yes 2 ☐ No | Searc 1 ☐
If yes, what was found: 1 ☐ Drugs 2 ☐ Drug Paraphernalia 3 ☐ Alcohol 4
If a search of the Vehicle was conducted, was contraband found? 1 ☐ Yes 2 ☐ No
If the contraband found was drugs, what was the amount? 1 ☐ < 2 grams 2 ☐ 2-10 grams
Driver | Consent Search Requested? 1 ☐ Yes 2 ☐ No | Consent Given? 1 ☐ Yes 2 ☐ No | Searc 1 ☐
Passenger(s) | Consent Search Requested? 1 ☐ Yes 2 ☐ No | Consent Given? 1 ☐ Yes 2 ☐ No | Searc 1 ☐
If a search of the Driver or Passenger(s) was conducted, was contraband found? 1 ☐ Yes
If yes, what was found: 1 ☐ Drugs 2 ☐ Drug Paraphernalia 3 ☐ Alcohol 4
If the contraband found was drugs, what was the amount? 1 ☐ < 2 grams 2 ☐ 2-10 grams

**Section C - Police Dog Sniff Sea

Illinois Traffic Stop Data - Findings

- **The analysis found that many police departments subject non-White drivers to significantly higher rates of dog sniffs and search requests than White drivers (filled in circles)**
- Statistical significance was calculated using a standard z-test for two population proportions (aka chi square test)

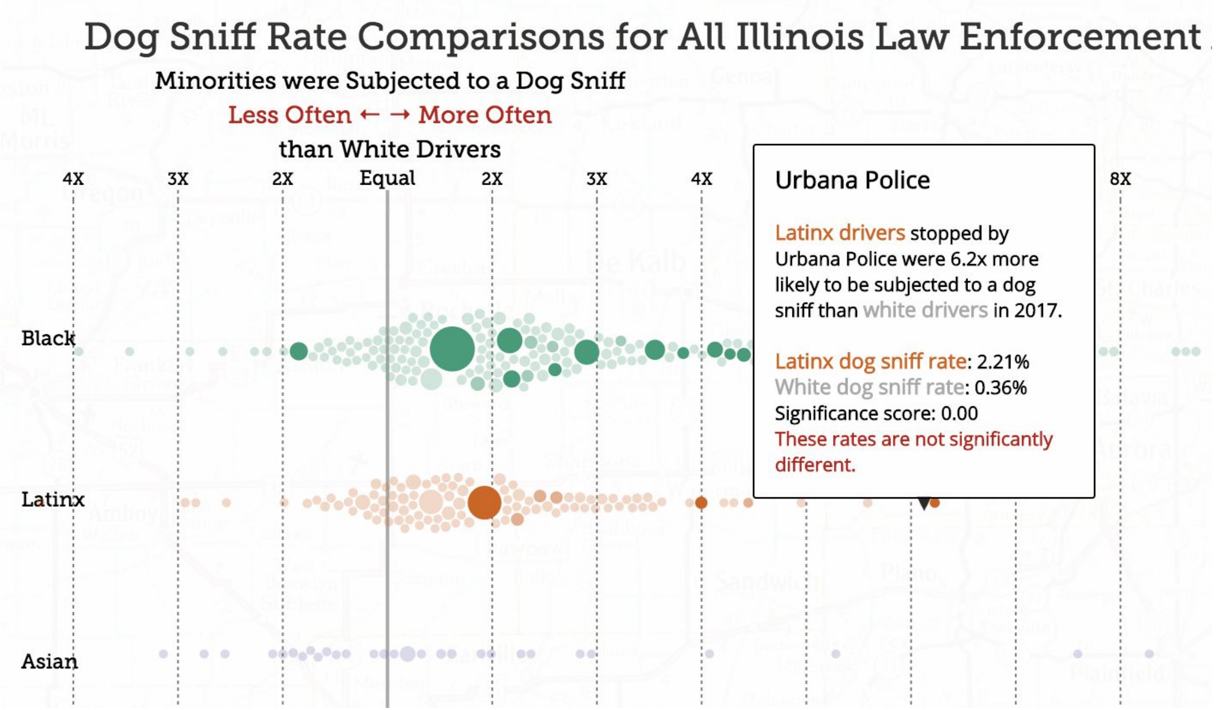


Image sourced from https://illinoistrafficstops.com/

Illinois Traffic Stop Data - Small n-sizes

- Police departments in rural areas do not stop as many non-White drivers simply because the population is skewed White
- Take Urbana Police: although Latinx drivers appear to be searched more often than White drivers, **we can't say that the difference is statistically significant because the standard error on our estimate is very large**
- In addition, small n-sizes often invalidate the assumptions of chi square tests

Urbana Police, 2017 Stop Data

	White Drivers	Latinx Drivers
Number of Dog Sniffs	6	3
Number of stops	1676	136
Dog Sniff Rate	0.0036	0.0221

Latinx Rate - White Rate = *0.0185*

Standard Error = *0.0288*

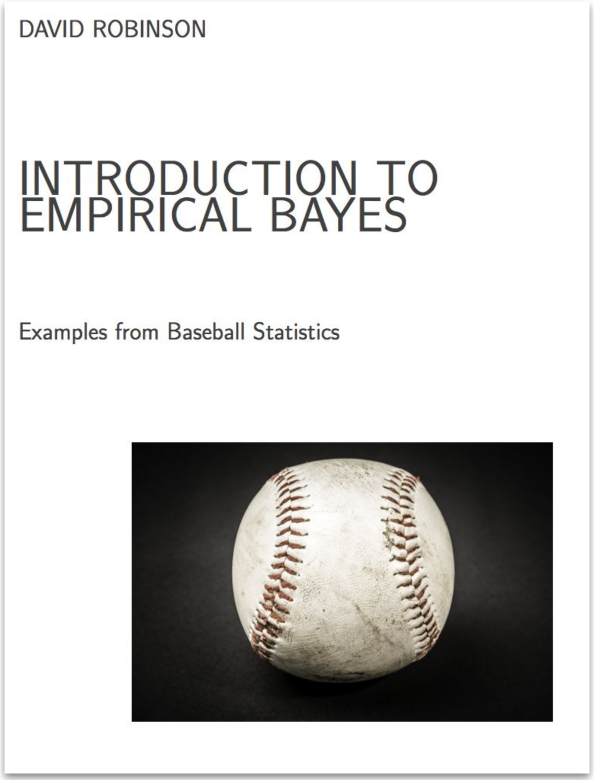
Confidence Interval = *(-0.0103, 0.0473)*

Data sourced from https://github.com/JustDSOrg/itssutils

How do we solve the small n-size problem?

Answer: Empirical Bayesian techniques!

- "**Empirical Bayes** methods are procedures for statistical inference in which the prior distribution is estimated from the data."
  - In other words, we can use information about ***all*** police departments to help us calculate more accurate estimates for ***individual*** departments
- Framework in this presentation comes from David Robinson's series of blog posts and book (@drob on Twitter)
- An example from baseball...who is the better hitter?
  - **Batter A:** 4 hits out of 10 = .400
  - **Batter B:** 100 hits out of 300 = .333



Answer: Probably Batter B!

- You can think of each person's true batting average as a distribution of possible values...the more information we know, the narrower the distribution will be

- Although Batter A seems better than Batter B, we have less information about them...they could've just gotten lucky! Therefore, their raw distribution of possible values is very wide.

- One thing we can do is look at the batting averages of ***all*** baseball players and then use that information as a *base*, or ***prior***, distribution of how we should judge someone with little to no information

Batting Averages of Thousands of MLB Players

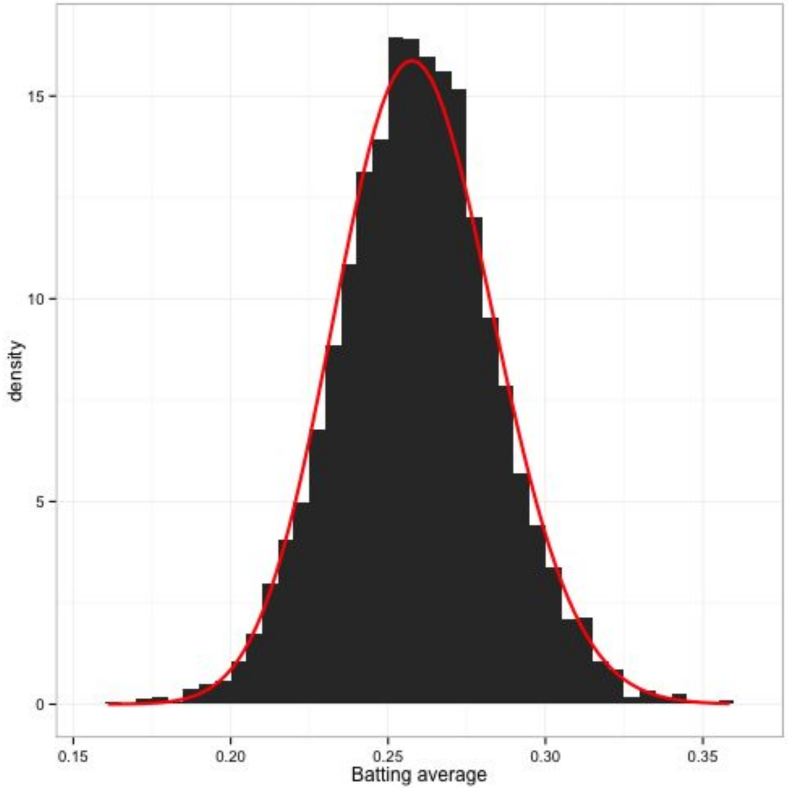
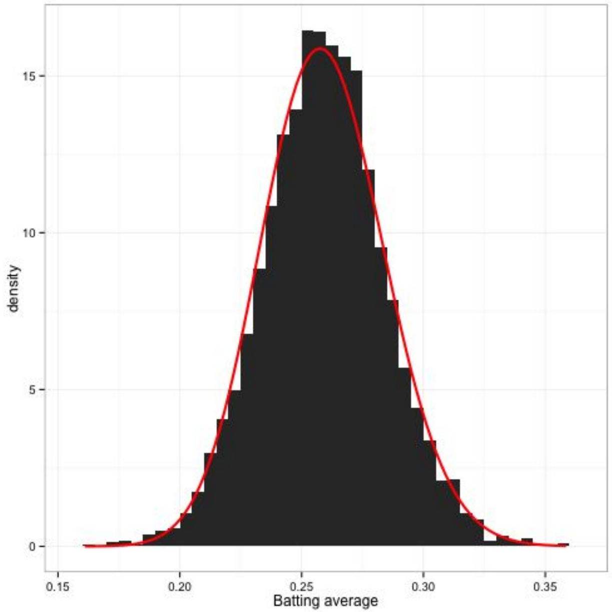


Image sourced from http://varianceexplained.org

Using a prior distribution gives us more confidence

Prior Distribution

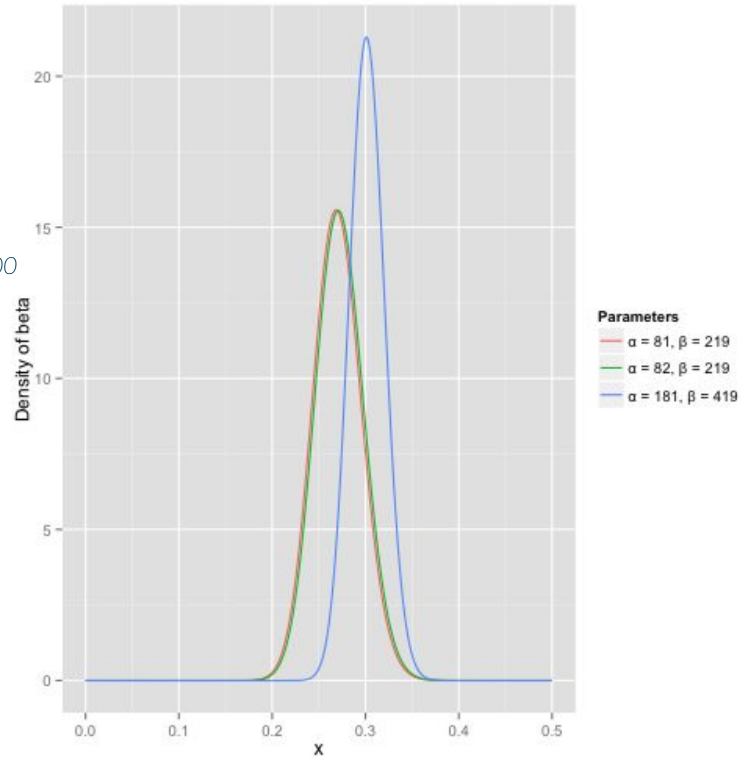


Red Line = *Prior*

Green Line = *Player with 1 hit out of 1*

Blue Line = *Player with 100 hits out of 300*

Posterior Distributions



Images sourced from http://varianceexplained.org

Empirical Bayes - Overview

Step 1: Estimate Prior Distribution	**Step 2:** Create Posterior Distributions	**Step 3:** Compare Resulting Distributions
Take the data and fit a distribution to it using maximum likelihood estimation	Use that information as a base to get individual distributions for each data point	Compare the resulting distributions to each other to make inferences
"What should be my expectation of police searches be with no other info?"	*"How should adjust my expectations given data about a particular department?"*	*"How do white and black search rates compare after we've adjusted our data?"*
ebbr::ebb_fit_prior()	**ebbr::add_ebb_estimate()**	**ebbr::add_ebb_prop_test()**

Empirical Bayes - Step 0: Get data

- Continue looking at Dog Sniff Rate = # Dog Sniffs / # Stops
- Each row is a unique Department / Year / Race combo
- Also limiting to rows with enough stops and sniffs; exclude rows where rate is zero for both races
- **Results in ~3.6K rows**

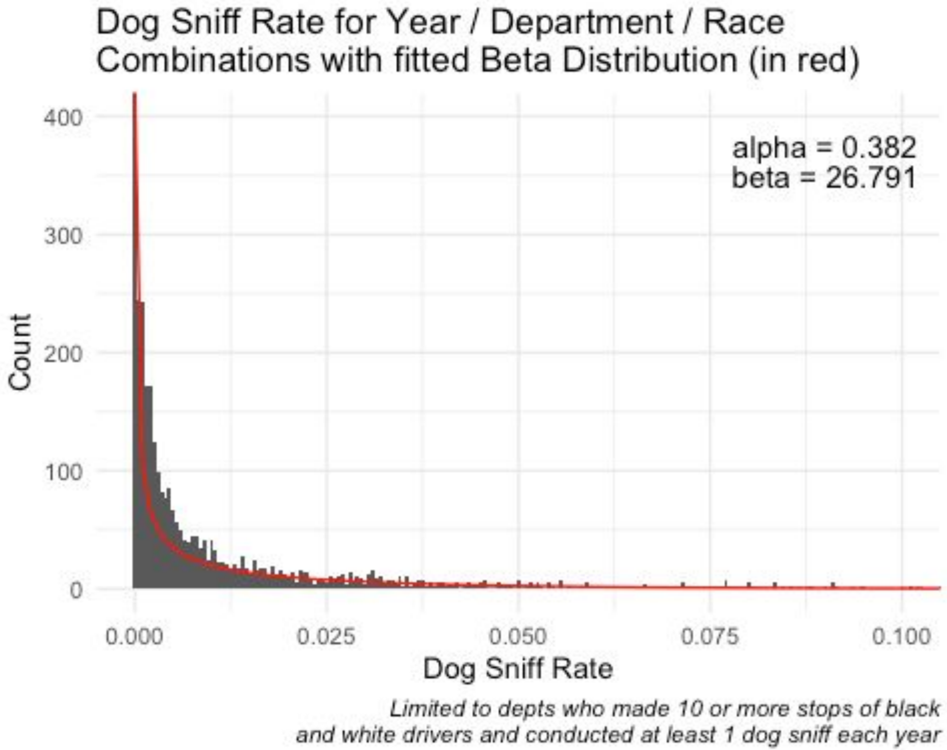
```
dog_sniff <-
  raw %>%
  select(AgencyName, Year, DriverRace,
         DogSniffCount, StopCount, DogSniffRate) %>%
  filter(Year != "All_Year",
         AgencyName != "All_AgencyName",
         DriverRace %in% c("Black", "White"),
         StopCount >= 10,
         !is.na(DogSniffCount))
```

```
## # A tibble: 6 x 6
##   AgencyName           Year  DriverRace DogSniffCount StopCount DogSniffRate
##   <chr>                <chr> <chr>              <dbl>     <dbl>        <dbl>
## 1 Adams County Sheri… 2012  Black                  2        15      0.133
## 2 Adams County Sheri… 2012  White                  4       422      0.00948
## 3 Adams County Sheri… 2013  Black                  1        22      0.0455
## 4 Adams County Sheri… 2013  White                  0       411      0
## 5 Adams County Sheri… 2015  Black                  2        44      0.0455
## 6 Adams County Sheri… 2015  White                  4       655      0.00611
```

Data sourced from https://github.com/JustDSOrg/itssutils

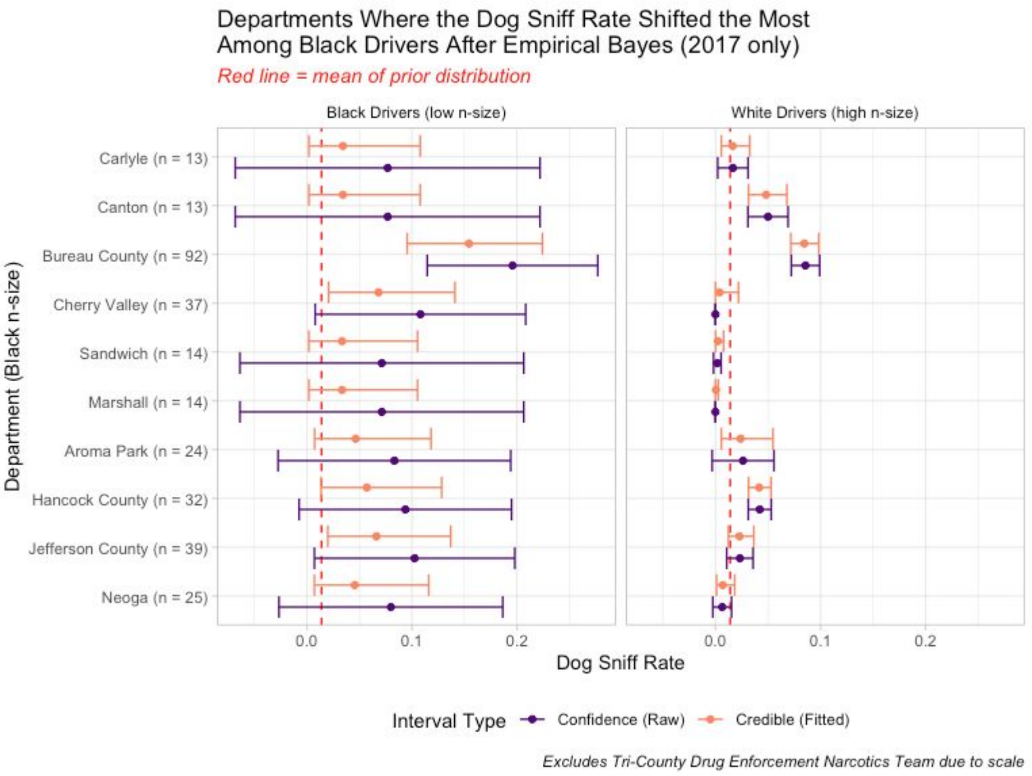
Empirical Bayes - Step 1: Fit prior distribution

- We fit a **beta distribution** to our data; it has very nice properties and values are always between 0 and 1
- Mean of distribution is $\alpha_0 / (\alpha_0 + \beta_0) = 0.014$
- Large alpha and beta parameters indicate more evidence, which allows us to have confidence in our initial distribution



Empirical Bayes - Step 2: Update data points

- We take the prior parameters and calculate posterior distributions (also a Beta dist.) for each row in our data
- "Fitted" mean = (α_0 + #Dog_Sniffs) / ($\alpha_0 + \beta_0$ + #Stops)
- *ebbr* does calculations for us and also gives **credible intervals** bounds
- Empirical Bayes "shrinks" the interval and moves the point estimate closer to the mean of the distribution



Empirical Bayes - Step 3: Compare

- For each department / year, take the new, fitted point estimates of dog sniffing and determine if the difference in the two races is significant
- Traditional statistics test (e.g. chi-square test) give a p-value, while Empirical Bayes gives a **posterior error probability (pep)**, which is different but analogous
- Look at Westville Police (2016) - we previously stated there was a difference, but with EB the difference is not significant

Westville Police, 2016 Stop Data

	Raw Data	Empirical Bayes
Black - Dog Sniff Rate	0.1818	0.0624
White - Dog Sniff Rate	0.0270	0.0254
Difference (B - W)	0.1548	0.0370
Hypothesis Test	*Is the difference in rates significant?*	
p-value (Raw) or pep (EB)	0.0320	0.2418

Empirical Bayes - Conclusions

- If we perform the same calculations for all departments, we see that Empirical Bayes significance tests agree with the original tests* for the vast majority of departments (96%)
- The ones which change often involve cases where there were low sample sizes or when one race did not see any dog sniffs at all
- If we had more data, we would anticipate switching more departments

Number of departments that either switched significance signs or stayed the same

Raw Data \ *Empirical Bayes*	Yes, statistical difference	No statistical difference
Yes, statistical difference	**349** *(20%)*	**61** *(3%)*
No statistical difference	**21** *(1%)*	**1359** *(76%)*

**For the purposes of this exercise, we are ignoring small sample size warnings of chi square test*
In practice, we would not estimate p-values for each of the departments

Caveats

1. This was a **very simplified** analysis, and there are many ways we could have improved it (accounted for rural vs urban, take overall police force size into account, etc.)
2. We could also apply this analysis to other rates in the data, but you need to make sure your data fits a beta distribution (our original data didn't fit perfectly)
3. Reducing the variance in your estimates comes at a cost of introducing bias
4. Many other contingency table tests we could've performed as well (e.g. Fisher's exact test)

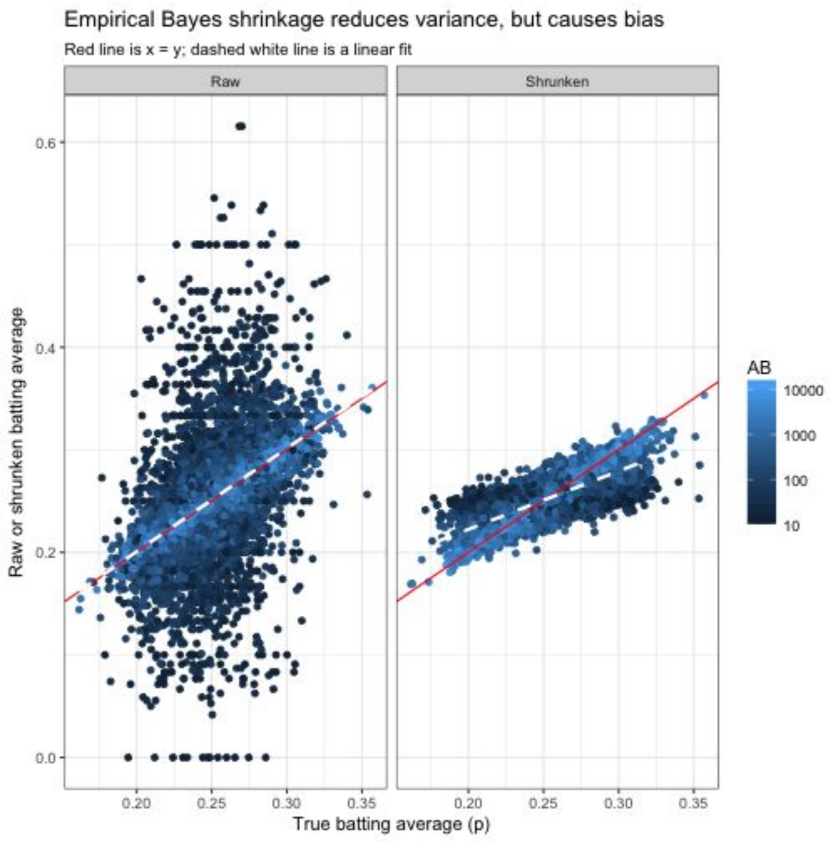


Image sourced from http://varianceexplained.org

Takeaways

- Empirical Bayes is a fairly powerful, yet simple, ways to improve estimates when you have small individual sample sizes, but collectively a lot of data
- The ebbr() package is capable of more advanced analyses such as hierarchical modeling and mixture modeling
- Outside of sports and traffic data, this type of analysis can be used for a variety of scenarios - medical research, advertising campaigns. If you have existing data, try it out!

Empirical Bayes showing how the advantage of left handed hitter has decreased over time

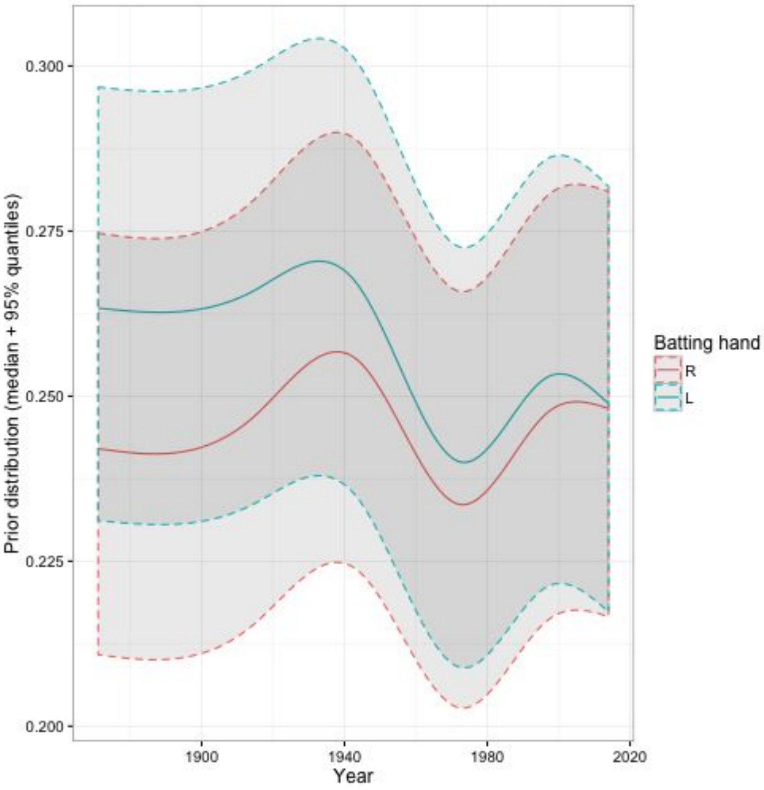


Image sourced from http://varianceexplained.org

Questions?