


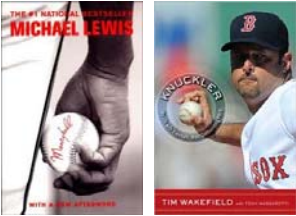


IT MAY BE A GREAT DAY FOR BASEBALL, BUT IS IT A GREAT DAY FOR A KNUCKLEBALL?

Robert Carver, Stonehill College
 2011 Joint Statistical Meetings Miami Beach, Florida
 Session 182 | August 1, 2011 | 10:30 AM
 rcarver@stonehill.edu




INTRODUCTION, CONTEXT, & DISCLAIMERS

- o Blessings/curses of using Baseball data
- o First foray into sabermetrics
- o Focus on *teaching* opportunities
- o Checking a hunch about a local icon...

SOME COMMON STAT ED REFRAINS

- o Real Data
- o Technology
- o Emphasize **concepts** like variation, its roots & impacts
- o Use engaging examples of data analysis
- o Connect to students' prior knowledge
- o Illustrate applications
- o Keep away from a Cookbooks.... no matter how much students like them



ANOTHER BOSTON-BASED ICON

wrote  and taught, taking her audience to heart...

with *real* ingredients, humor, enthusiasm...


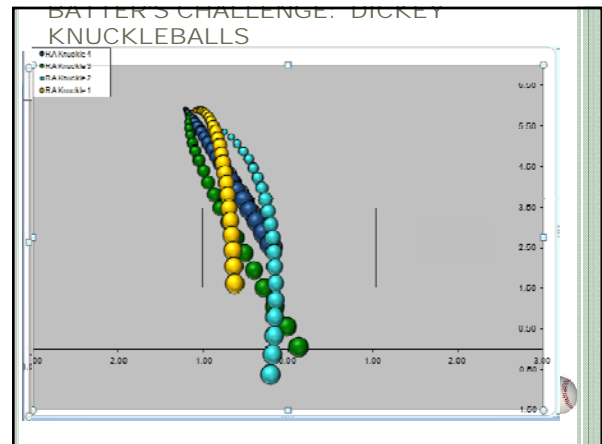
...and lots of preparation & help

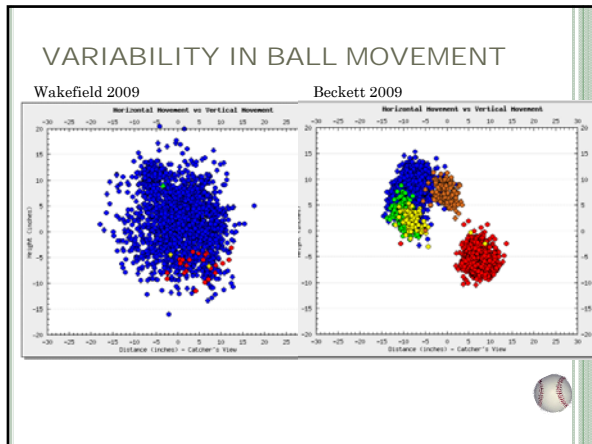





AN ALL-ENCOMPASSING DISCIPLINE DESERVES ALL-ENCOMPASSING EXAMPLES

- o Physics
- o Meteorology
- o Sports
- o Management
- o Game Theory
- o Economics
- o Digital Technology



WHAT CAUSES PITCH MOVEMENT?

- Amount of rotation imparted by pitcher
- Initial orientation of stitches
- Pitch velocity imparted
- Interaction of stitched seams, rotation, & air currents
 - Pressure differentials
 - Magnus Effect

KNUCKLEBALL

- Minimal rotation
- Low velocity
- Orientation of ball leaving hand
- Asymmetry of stitch pattern → unequal turbulence
- Unpredictable for catcher, batter, pitcher...

Plausible atmospheric influences:

- Air pressure
- Wind speed
- Wind direction
- Humidity
- Temperature

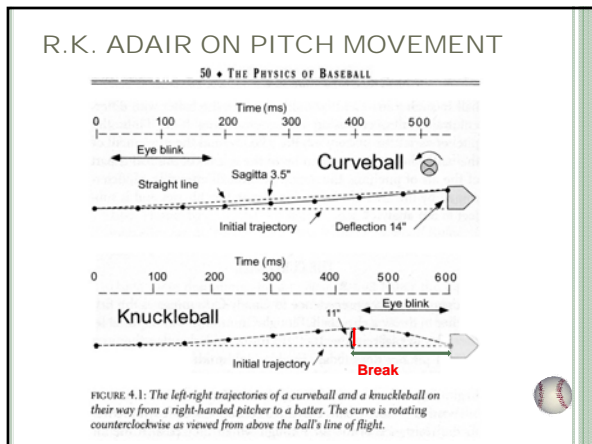
R.K. ADAIR ON PITCH MOVEMENT

“The disadvantage of the knuckleball ... is that the forces can vary strongly with very small differences in orientation of the ball; hence the pitcher, however skilled, finds it very difficult to control the pitch.

If it breaks sharply, it is difficult to catch and leads to too many passed balls.

If it doesn't break, it is no more than a batting practice pitch, which is even worse.”

The Physics of Baseball, 2002 (p. 55)



QUESTIONS

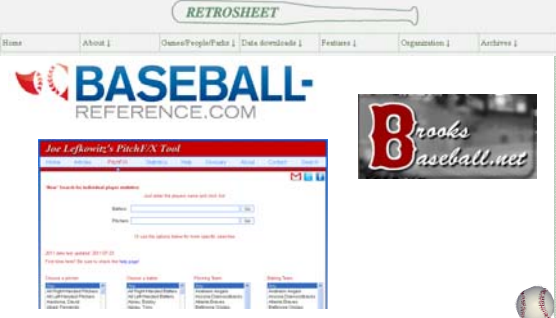

- How much, if at all, do atmospheric conditions affect knuckleball movement, after controlling for imparted spin & orientation?
- Can pre-game weather conditions predict pitching performance for a knuckleball pitcher?
 - How should we **measure performance**?

TEACHING OPPORTUNITIES



- Study Design
- Conceptual Foundations
- Data & Measurement Issues
- Data Acquisition
- Data Preparation
- Multivariate Modeling & Interpretation



GOOD NEWS: FREE SOURCES OF ABUNDANT BASEBALL DATA





NWS DATA INTERFACE: (SMALL CHARGE)


SOME STUDY DESIGN COMMENTS

- Observational data – class discussion re: DOE?
- Integration of data from three sources (too much for intro students to do, but worth pointing out)
 - Weather data from NWS and Retrosheet diverge
 - This analysis restricted weather observations to game time
- Unit of Analysis: Pitch/at-bat/game...
- Limited observations to Fenway Park to control for:
 - Ballpark orientation
 - Altitude
 - Effects of walls & stadium structure on wind
- Only analyzed performance of Starting Pitchers




TWO INITIAL MODELS

Mean Break Length <ul style="list-style-type: none"> ○ Mean (Spin Axis) ○ Std Dev (Spin Axis) ○ Mean (Spin Rate) ○ Wind Speed ○ Crosswind Dummy ○ DewPoint ○ Temperature ○ Barometric Pressure 	Game ERA for Starter <ul style="list-style-type: none"> ○ Wind Speed ○ Wind Out Dummy ○ Relative Humidity ○ Temperature ○ Barometric Pressure ○ Days Rest ○ Opp Team Batting Avg
---	--




BALL MOVEMENT – MEAN BREAK LENGTH

Predictor	Wakefield KN 2008-10; n=22 games
Intercept	NS
Mean (Spin Dir)	+ **
Std Dev (Spin Dir)	+ ***
Mean (Spin Rate)	NS ~
Wind Speed (MPH)	NS
DewPoint	NS ~
Temp (Fahrenheit)	NS
Sea Level Pressure	NS
Cross Wind	+ ***
R-sqr Adj. R-sqr	0.78 0.66
F	6.017 ***




REFINED MODEL OF BREAK LENGTH


Predictor	Wakefield KN 2009-10; n=22 games
Intercept	10.32 ***
Mean(Spin Dir)	- 0.023 ***
StdDev(Spin Dir)	0.014 **
Mean (Spin Rate)	- 0.002 *
CrossWind	- 0.700 ***
DewPoint	- 0.012 *
R-sqr Adj. R-sqr	0.77 0.70
F	10.787 ***




ERA MODELS ESTIMATED COEFFICIENTS

Predictor	Wakefield (n=40)	Rest of Staff (n=203)
Intercept		
Opp Team Batting Avg	-25.26 **	
Days Rest		0.015 **
Wind Speed (mph)	- 0.07 ***	
Wind Out (0-1)	- 0.42 **	
Relative Humidity		
Sea Level Pressure	1.01 **	
R-sqr Adj. R-sqr	0.46 0.36	0.03 0.001
F	4.79 ***	1.049 (n.s.)





- ### DEPENDENT VARIABLE: BILL JAMES' GAME SCORE
- Incorporates "pitcher-specific" outcomes
 - Higher Game Score → better performance
 - Wakefield only: 2008-10
 - With same variables as ERA model, poor results. WINDOUT dummy is significant
 - If we add Mean(Break) to the model:
 - Break Len coeff is +9.9
 - Windout coeff is +13.6
 - Both signif at 0.05
 - R² is 0.29
- 

- ### NEXT STEPS
- Apply these models to 2011 game data & check predictive validity
 - Examine pitch-by-pitch movement data
 - How to integrate NWS data usefully?
 - Dew points relatively stable for duration
 - Wind direction and speed much more variable
 - Further examine "pitcher-only" measures of effectiveness
 - Further examine relationship of pitch movement to outcomes
- 

TAKE AWAY MESSAGES

- Knuckleball movement = f(pitcher skill, wind direction, dewpoint)
- Effectiveness = f(movement, offense, wind direction, wind speed)

In short: at least part of the answer, my friends, is blowin' in the wind...

- ### QUESTIONS?
- Acknowledgement:
 - Jim Albert, Bowling Green
 - Contact Information:
 - rcarver@stonehill.edu
- 