

# **Tufts University EX-13 The Analysis of Baseball: Statistics and Sabermetrics**

Spring 2005

## **Instructor:**

**Andy Andres, PhD, N'99**  
Assistant Professor of Natural Science, Boston University  
Harvard College Tutor in Biology  
Data Analyst for BaseballHQ.com

## **Teaching Assistants:**

**Morgan Melchiorre**  
**Defensive Wizard Guy and The Guy Who Takes Care Of All Things That  
Need To Get Done**  
Information Systems Specialist, Jean Mayer Human Nutrition Research Center on  
Aging at Tufts University

**David Tybor, MS, N'03; MPH, M'03**  
**Stats Guy**  
Doctoral Student in Nutritional Epidemiology, Friedman School of Nutrition

**Web Site:** [sabermetrics.hnrc.tufts.edu](http://sabermetrics.hnrc.tufts.edu)  
**Email Contact:** [Jumbo.SABR@gmail.com](mailto:Jumbo.SABR@gmail.com)  
**Time and Place:** **Thursdays, 5:30 – 8:15, Miner 112**  
**Office Hours:** **Thursdays, 3:30 – 6:00, Location TBD**

## **COURSE DESCRIPTION**

This course will teach Tufts students the fundamentals of Sabermetrics, the analysis of baseball through objective evidence. It will cover the important concepts in statistics needed to perform sabermetric research and analysis. Students will design and implement their own sabermetric research study. We will discuss baseball not through conventional wisdom and consensus, but by searching for the truth of baseball performance using baseball statistics. Hitting performance, pitching performance, defensive performance will all be analyzed and better understood by looking at and analyzing current and historical data from baseball.

## **GOALS FOR THE COURSE**

Students at the Ex College taking this course will learn about the fundamentals of baseball analysis such that they will be able to research, design, and implement their own sabermetric studies.

# **SYLLABUS**

## **1/20/05: Class 1: Introduction to Sabermetrics, the Statistics of Baseball**

Introduction to Sabermetrics. Sabermetric Basics. Course Outline. Case Study 1: the MVP of the World Series. Case Study 2: Being Jose Oquendo. Case Study 3: Give Jeter Some Love.

Presentation of Student Research.

### **Reading for Next Class:**

Moneyball

## **1/27/05: Class 2: Hitting, Part I**

Student Introductions. Discussion of Moneyball. The Scientific Method in Sabermetrics.

Introduction to Hitting Quantification.

Introduction to Statistics for Sabermetric Research.

### **Reading for Next Class:**

Norusis, Counting Responses; Computing Descriptive Statistics  
Hitting Reading Packet, distributed in class

### **Homework 1:**

Critique, in one typed page, an observation on baseball from the recent popular press. Use this observation to generate a hypothesis or model. Explain briefly what sort of information you would need to test your model. Describe how the information you gather might affect your model. For example, the Boston Globe once noted, sans context, that Jason Varitek was batting .375 (36 hits in 96 at-bats) after a day off. An objective fan might ask, "Why is this writer telling me this? Is this performance from Varitek considerably higher than his usual performance? If so, do other players on the this year's team hit higher after an off-day? Do other catchers in the league hit higher after a day of rest? What did Varitek hit after a day off last year, etc?" Such thoughts can be used to generate a model with an appropriate hypothesis, and a research method to analyze them. Be sure to cite your source, which may be a newspaper, magazine, website, or television broadcast. Creativity counts. ***Due in class next week.***

*This Week in Baseball History (2/2/1876): The National League is officially formed with teams located in Boston, Chicago, Cincinnati, Hartford, Louisville, New York, Philadelphia and St. Louis.*

## 2/3/05: Class 3: Hitting, Part II

Limitations of Traditional Batting Statistics. Count vs. Rate Statistics. Details of OPS, OPS+, and Advanced Methods of Rating the Importance of Hitting Statistics. Runs Created.

Lab Session: Introduction to SPSS.

### **Reading for Next Class:**

DIPS 2004: <http://www.futilityinfielder.com/dips04.html>  
James, 885-8 DIPS; James, 289-94 Mark Fidrych and the K  
Norusis, Linear Regression and Correlation  
Albert, Chapter 4, Relationships Between Measurement Variables

### **Homework 2:**

SPSS Analysis, to be distributed in lab session. *Due in class next week.*

*This Week in Baseball History (2/7/1979): Jesse Orosco becomes 'the player to be named later' in the Mets' trade of Jerry Koosman to the Twins.*

## 2/10/05: Class 4: Pitching

Quantification of Pitching. Charting and Predicting. Limitations of Traditional Pitching Statistics. Defense Independent Pitching Statistics (DIPS).

### **Reading for Next Class:**

Norusis, Chapters 12, 13, 14, Testing Hypotheses About Means  
Albert Chapter 3, Comparing Batches and Standardization  
Case Study: Mazzone Effect

### **Homework 3:**

**Project:** Describe three potential topics for your research project. In one or two typed pages, state the seminal observation, research hypothesis, rough methodology for testing, and potential data sources, if applicable. *Document is due in class next week.*

*This Week in Baseball (2/15/2005): Washington Nationals' Pitchers and Catchers Report, Melbourne, FL*

## **2/17/05: Class 5: Using Statistics in Sabermetrics**

SPSS Homework Review. Correlation. Linear Regression. Regression to the Mean. Statistical Significance. Testing Hypotheses About Means, Case Study: The Mazzone Effect.

### **Reading for Next Class:**

James 439 Gold Gloves; 456 Buckner; 460 Making Outs?; 487-90 Nap; 495 Double Play; 501-5 Maz, Frank White.  
Defensive Reading Packet, distributed in class  
Norusis, Comparing Observed and Expected Counts

### **Homework 4:**

**Project:** Meet individually with instructors *before next scheduled course* to discuss research proposals. After meeting, begin research. Background section of paper – a thorough description of what is known about your topic – is due **3/3/05**.

*This Week in Baseball (2/24/2005): First full workout, Oakland Athletics, Phoenix, AZ*

## **2/24/05: No Class. Individual meetings 2/24 or 2/25**

*This Week in Baseball (3/02/2005): Spring Training Opener: NYM @ WAS; COL @ CWS*

## **3/3/05: Class 6: Defense**

Shortcomings of Traditional Fielding Statistics. Exploring Range Factor, Zone Rating, and Advanced Metrics. Catcher's ERA.

Comparing Observed and Expected Counts Using Chi-Square.

### **Reading for Next Class:**

James, 330-371 Win Shares; 422 Mike Macfarlane  
“Putting it All Together” Reading Packet, distributed in class

### **Homework 5:**

Short writing assignment, to be distributed in class. ***Due in class next week.***

*This Week in Baseball History (3/05/1973): Yankee teammates Fritz Peterson and Mike Kekich announced they have traded families – including their wives, children and dogs.*

### **3/10/05: Class 7: Putting It All Together**

Overall Summary Statistics: Value Over Replacement Player (VORP), Bill James' Win Shares, Runs Above Replacement (RAR). Feedback on Project Ideas.

#### **Reading for Next Class:**

“Introducing PECOTA,” Nate Silver, *Baseball Prospectus* 2003, p. 507-514  
Norusis, Testing Regression Hypotheses  
Norusis, Building Multiple Regression Models  
James, 480, 715 Brains; 515-6 Trading Talent.  
“Fanalytics,” Ron Shandler, *Baseball Forecaster* 2005

#### **Homework 6:**

Based on background research and feedback on project proposals, identify appropriate data sources for analysis. Provide a typed description of said data, including, if appropriate, source, format, size, program to be used, etc. *Due in class next week.*

*This Week in Baseball History (3/12/2001): Richard Hidalgo signs the third-richest contract in Astro history with a four-year \$32 million deal with the club.*

### **3/17/05: Class 8: Projection**

Predictions with Multiple Linear Regression Models. PECOTA and Advanced Projection.

#### **Reading for Next Class:**

Evaluation Reading Packet, distributed in class

#### **Homework:**

Work on project over spring break. Be prepared to discuss successes, preliminary findings, and barriers during next class.

*This Week in Baseball History (3/22/1908): Tigers' outfielder Ty Cobb signs for \$4000 with an \$800 bonus if he hits over .300.*

### **3/31/05: Class 9: Evaluation**

Putting a Value on MLB Players. Student Project Roundtable.

#### **Homework:**

Continue work on project.

*This Week in Baseball (4/3/2005): Opening Day, BOS @ NYY*

#### **4/7/05: Class 10: Assorted Topics in Sabermetrics**

The Field Manager. Performance Enhancing Substances. The History of Sabermetrics.

*This Week in Baseball (4/11/2005): Fenway Home Opener, NYY @ BOS*

#### **4/14/05: Class 11: Student Presentations**

Student Presentations. Schedule TBD.

#### **Homework:**

Final *papers* for all students due next week.

*This Week in Baseball 4/18/2005: Patriot's Day, Boston Marathon, TOR @ BOS, 11:05 AM*

#### **4/21/05: Class 12: Student Presentations**

Student Presentations. Schedule TBD.

Student Papers Distributed for Peer Review. Review due next week.

#### **4/28/05: Class 13: Guest Speaker and Course Evaluation**

---

#### **BOOKS:**

Moneyball by Michael Lewis

SPSS 11.0 Guide to Data Analysis by Marija Norusis

The New Historical Baseball Abstract by Bill James

Recommended: Total Baseball by John Thorn

Various Online Content Sources for Baseball Statistics and Analysis:

Baseballprospectus.com

Stats, Inc. (stats.com)

BaseballHQ.com

Espn.com

#### **EVALUATION**

Participation in Seminar (including Peer Review of Research): 30%

Homework Assignments: 30%

Research Paper: 30%

Final Presentation: 10%