ORF 307: Lecture 8

Academics and Athletics

Robert Vanderbei

February 28, 2019

Slides last edited on January 23, 2019

Fair Grading

Senior Thesis of Gordon Scharf'09

Beatles University

	MAT	CHE	ANT	REL	POL	ECO
John	B−	B	B+	A−		
Paul	C+	B−		B+	A−	
George		C+	B−		B+	A−
Ringo			C+	B−	B	B+

where

MAT 402: Advanced Addition

CHE 333: Intermediate Explosives

ANT 536: The Behavior of Anthropologists

REL 101: Atheism

POL 242: Constitutional Manipulation

ECO 666: The Root Of All Evil

The Model

We assume that every grade, $g_{i,j}$ for student i in course j, can be decomposed as a sum

1. *aptitude*, a_i, of student i,
2. *easiness*, e_j, of course j,
3. plus some (presumably small) error $\epsilon_{i,j}$.

That is,

$$g_{i,j} = a_i + e_j + \epsilon_{i,j}.$$

The $g_{i,j}$'s are data. We wish to find the a_i's and the e_j's that minimize the sum of the absolute values of the $\epsilon_{i,j}$'s:

$$\begin{aligned} &\text{minimize} && \sum_{i,j} |\epsilon_{i,j}| \\ &\text{subject to} && g_{i,j} = a_i + e_j + \epsilon_{i,j} \qquad \text{for all students/course pairs } (i,j) \\ & && \sum_j e_j = 0. \end{aligned}$$

Absolute Value Trick Again

$$\begin{aligned}
\text{minimize} \quad & \sum_{i,j} |\epsilon_{i,j}| \\
\text{subject to} \quad & g_{i,j} = a_i + e_j + \epsilon_{i,j} \qquad \text{for all students/course pairs } (i,j) \\
& \sum_j e_j = 0.
\end{aligned}$$

is equivalent to

$$\begin{aligned}
\text{minimize} \quad & \sum_{i,j} t_{i,j} \\
\text{subject to} \quad & g_{i,j} - a_i - e_j \leq t_{i,j} \qquad \text{for all students/course pairs } (i,j) \\
& -t_{i,j} \leq g_{i,j} - a_i - e_j \qquad \text{for all students/course pairs } (i,j) \\
& \sum_j e_j = 0.
\end{aligned}$$

The AMPL Model

```
set STUDENTS;
set COURSES;
set GRADES within {STUDENTS, COURSES};

param grade {GRADES};

var aptitude {STUDENTS};
var easiness {COURSES};
var dev {GRADES} >= 0;

minimize sum_dev: sum {(s,c) in GRADES} dev[s,c];

subject to def_pos_dev {(s,c) in GRADES}:   aptitude[s] + easiness[c] - grade[s,c] <= dev[s,c];

subject to def_neg_dev {(s,c) in GRADES}:   -dev[s,c] <= aptitude[s] + easiness[c] - grade[s,c];

subject to normalized_easiness:   sum {c in COURSES} easiness[c] = 0;

data;
set STUDENTS := include "names.txt" ;
set COURSES := include "courses.txt" ;
param: GRADES: grade := include "namecoursegrade.txt" ;

solve;
```

Model and three data files are here...

https://vanderbei.princeton.edu/307/models/grades/ratings.txt
https://vanderbei.princeton.edu/307/models/grades/names.txt
https://vanderbei.princeton.edu/307/models/grades/courses.txt
https://vanderbei.princeton.edu/307/models/grades/namecoursegrade.txt

Data from HARVARD's School of Science

Alex	Statistics	B-
Alex	Epistemology	B-
Alex	Phrenology	A-
Alex	Pharmacology	B
Alex	Philology	A-
Alex	Theology	A+
Alex	Geology	B+
Alex	Gynecology	A
Andy	Scientology	B-
Andy	Etymology	B-
Andy	Sociology	A-
Andy	Psychology	B
Andy	Cosmology	A-
Andy	Eulogy	A+
Andy	Immunology	B+
Andy	Methodology	A
Ariel	Epistemology	B
Ariel	Topology	B+
Ariel	Etymology	A+
Ariel	Eulogy	A+
Ariel	Genealogy	A-
Ariel	Morphology	A+
Ariel	Pathology	A-
Ariel	Technology	A+
Billy	Etymology	A+
Billy	Cosmology	A+
Billy	Philology	B
Billy	Ideology	A+
Billy	Immunology	A+
Billy	Methodology	A+
Bobby	Topology	A+
Bobby	Philology	B+
Bobby	Eulogy	A+
Bobby	Genealogy	A
Bobby	Pathology	A+
Bobby	Urology	A+
Brett	Topology	A+
Brett	Pharmacology	A-
Brett	Demonology	A+
Brett	Ecology	A+
Brett	Ideology	A+
Brett	Ontology	A
Brett	Technology	A+
Brook	Topology	A
Brook	Phrenology	B-
Brook	Cosmology	A
Brook	Philology	C+
Brook	Theology	B
Brook	Demonology	A-
Brook	Technology	A-
Brook	Terminology	A+
Cameron	Etymology	A
Cameron	Demonology	A
Cameron	Eulogy	A
Cameron	Ecology	B+
Cameron	Ontology	B
Cameron	Pathology	A
Cameron	Tautology	B+
Cameron	Technology	A-
Cary	Psychology	A+
Cary	Philology	A
Cary	Eulogy	A+
Cary	Methodology	A+
Cary	Morphology	A+
Cary	Tautology	A+
Cary	Urology	A+
Casey	Statistics	A-
Casey	Scientology	A+
Casey	Eulogy	A+
Casey	Geology	B+
Casey	Immunology	A+
Casey	Methodology	A+
Casey	Technology	A+
Casey	Terminology	A+
Chris	Topology	A+
Chris	Phrenology	A
Chris	Pharmacology	A

For full list of grades, click here.

Student-by-Student Output

```
Alex (4.18)
 2.7   3.4    Statistics (-0.8)
 2.7   3.3    Epistemology (-0.9)
 3.7   3.7    Phrenology (-0.5)
 3.0   3.7    Pharmacology (-0.5)
 3.7   3.6    Philology (-0.6)
 4.3   3.6    Theology (-0.6)
 3.3   3.3    Geology (-0.9)
 4.0   3.7    Gynecology (-0.5)
 3.43  3.54   -- avg -- (-0.64)
---------------------------------
Andy (3.02)
 2.7   3.4    Scientology (0.4)
 2.7   3.4    Etymology (0.4)
 3.7   3.4    Sociology (0.4)
 3.0   3.4    Psychology (0.4)
 3.7   3.4    Cosmology (0.4)
 4.3   3.4    Eulogy (0.4)
 3.3   3.4    Immunology (0.4)
 4.0   3.4    Methodology (0.4)
 3.43  3.44   -- avg -- ( 0.42)
---------------------------------
Ariel (3.88)
 3.0   3.0    Epistemology (-0.9)
 3.3   4.0    Topology (0.1)
 4.3   4.3    Etymology (0.4)
 4.3   4.3    Eulogy (0.4)
 3.7   4.0    Genealogy (0.1)
 4.3   4.3    Morphology (0.4)
 3.7   4.3    Pathology (0.4)
 4.3   4.0    Technology (0.1)
 3.86  4.02   -- avg -- ( 0.15)
---------------------------------
```

```
Billy (3.88)
 4.3   4.3    Etymology (0.4)
 4.3   4.3    Psychology (0.4)
 4.3   3.7    Apology (-0.2)
 4.3   4.3    Cosmology (0.4)
 3.0   3.3    Philology (-0.6)
 4.3   4.0    Ideology (0.1)
 4.3   4.3    Immunology (0.4)
 4.3   4.3    Methodology (0.4)
 4.14  4.06   -- avg -- ( 0.19)
---------------------------------
Bobby (3.88)
 4.3   4.0    Topology (0.1)
 3.3   3.3    Philology (-0.6)
 4.3   4.3    Eulogy (0.4)
 4.0   4.0    Genealogy (0.1)
 4.3   4.3    Pathology (0.4)
 4.3   4.3    Urology (0.4)
 4.08  4.03   -- avg -- ( 0.16)
---------------------------------
Brett (4.18)
 4.3   4.3    Topology (0.1)
 3.7   3.7    Pharmacology (-0.5)
 4.3   4.3    Demonology (0.1)
 4.3   4.0    Ecology (-0.2)
 4.3   4.3    Ideology (0.1)
 4.0   4.0    Ontology (-0.2)
 4.3   4.3    Technology (0.1)
 4.17  4.13   -- avg -- (-0.05)
---------------------------------
```

```
Brook (3.58)
 4.0   3.7    Topology (0.1)
 2.7   3.1    Phrenology (-0.5)
 4.0   4.0    Cosmology (0.4)
 2.3   3.0    Philology (-0.6)
 3.0   3.0    Theology (-0.6)
 3.7   3.7    Demonology (0.1)
 3.7   3.7    Technology (0.1)
 4.3   4.0    Terminology (0.4)
 3.46  3.52   -- avg -- (-0.05)
---------------------------------
Cameron (3.58)
 4.0   4.0    Etymology (0.4)
 4.0   3.7    Demonology (0.1)
 4.0   4.0    Eulogy (0.4)
 3.3   3.4    Ecology (-0.2)
 3.0   3.4    Ontology (-0.2)
 4.0   4.0    Pathology (0.4)
 3.3   3.7    Tautology (0.1)
 3.7   3.7    Technology (0.1)
 3.66  3.74   -- avg -- ( 0.16)
---------------------------------
Cary (3.88)
 4.3   4.3    Psychology (0.4)
 4.0   3.3    Philology (-0.6)
 4.3   4.3    Eulogy (0.4)
 4.3   4.3    Methodology (0.4)
 4.3   4.3    Morphology (0.4)
 4.3   4.0    Tautology (0.1)
 4.3   4.3    Urology (0.4)
 4.26  4.11   -- avg -- ( 0.24)
---------------------------------
```

Full set of output is here.

Course-by-Course Output

```
Statistics (-0.76)
 2.7   3.4    Alex (4.2)
 3.7   3.1    Casey (3.9)
 3.7   3.1    Dale (3.9)
 3.0   3.1    Darcy (3.9)
 2.7   3.1    Emerson (3.9)
 2.3   2.5    Esme (3.3)
 2.7   3.1    Harley (3.9)
 2.3   2.4    Jordan (3.2)
 4.0   3.4    Max (4.2)
 3.7   3.4    Porntip (4.2)
 2.3   2.2    Sunny (3.0)
 3.3   3.1    Sydney (3.9)
 3.03  3.01   -- avg -- ( 3.77)
---------------------------------
Epistemology (-0.88)
 2.7   3.3    Alex (4.2)
 3.0   3.0    Ariel (3.9)
 2.7   2.7    Jade (3.6)
 2.0   2.4    Lou (3.3)
 4.0   3.3    Max (4.2)
 2.3   2.7    Robin (3.6)
 2.7   2.0    Sam (2.9)
 2.3   2.3    Tracy (3.2)
 2.71  2.71   -- avg -- ( 3.59)
---------------------------------
Scientology (0.42)
 2.7   3.4    Andy (3.0)
 4.3   4.3    Casey (3.9)
 4.3   4.3    Dale (3.9)
 4.3   4.3    Dominique (3.9)
 3.7   3.7    Esme (3.3)
 4.3   4.5    Lindsey (4.1)
 4.3   4.3    Meryl (3.9)
 4.3   4.0    Robin (3.6)
 4.3   4.3    Sydney (3.9)
 4.06  4.13   -- avg -- ( 3.70)
---------------------------------
Astrology (0.12)
 4.0   4.0    Darcy (3.9)
 4.3   4.3    Daryl (4.2)
 3.7   3.7    Jade (3.6)
 4.3   4.1    Morgan (4.0)
 3.0   3.1    Sunny (3.0)
 3.86  3.85   -- avg -- ( 3.73)
---------------------------------
Topology (0.12)
 3.3   4.0    Ariel (3.9)
 4.3   4.0    Bobby (3.9)
 4.3   4.3    Brett (4.2)
 4.0   3.7    Brook (3.6)
 4.3   4.3    Chris (4.2)
 4.3   4.3    Drew (4.2)
 3.3   3.3    Jordan (3.2)
 4.0   4.0    Kelly (3.9)
 4.0   4.1    Morgan (4.0)
 4.3   4.0    Reese (3.9)
 4.3   4.3    Skye (4.2)
 4.3   4.0    Sydney (3.9)
 4.06  4.03   -- avg -- ( 3.91)
---------------------------------
Phrenology (-0.48)
 3.7   3.7    Alex (4.2)
 2.7   3.1    Brook (3.6)
 4.0   3.7    Chris (4.2)
 4.3   3.4    Darcy (3.9)
 3.0   3.0    Devyn (3.5)
 3.3   3.1    Jade (3.6)
 4.3   3.6    Lindsey (4.1)
 3.7   3.7    Max (4.2)
 3.3   3.4    Meryl (3.9)
 3.3   3.5    Morgan (4.0)
 2.0   2.5    Sunny (3.0)
 3.7   3.4    Tanner (3.9)
 3.44  3.35   -- avg -- ( 3.82)
---------------------------------
Pharmacology (-0.48)
 3.0   3.7    Alex (4.2)
 3.7   3.7    Brett (4.2)
 4.0   3.7    Chris (4.2)
 3.0   3.4    Kelly (3.9)
 3.7   3.7    Kim (4.2)
 4.3   3.7    Porntip (4.2)
 3.3   3.4    Sydney (3.9)
 3.57  3.61   -- avg -- ( 4.09)
---------------------------------
Etymology (0.42)
 2.7   3.4    Andy (3.0)
 4.3   4.3    Ariel (3.9)
 4.3   4.3    Billy (3.9)
 4.0   4.0    Cameron (3.6)
 3.0   3.7    Dara (3.3)
 3.7   4.3    Harley (3.9)
 3.7   3.7    Lou (3.3)
 4.3   4.0    Peyton (3.6)
```

For full list, click here.

Least Squares is Better

$$\begin{aligned}
&\text{minimize} && \sum_{i,j} \epsilon_{i,j}^2 \\
&\text{subject to} && g_{i,j} = a_i + e_j + \epsilon_{i,j} && \text{for all students/course pairs } (i,j) \\
& && \sum_j e_j = 0.
\end{aligned}$$

The AMPL Model

```
set STUDENTS;
set COURSES;
set GRADES within {STUDENTS, COURSES};

param grade {GRADES};

var aptitude {STUDENTS};
var easiness {COURSES};
var dev {GRADES};

minimize sum_dev: sum {(s,c) in GRADES} dev[s,c]^2;

subject to def_dev {(s,c) in GRADES}:   aptitude[s] + easiness[c] - grade[s,c] = dev[s,c];

subject to normalized_easiness:   sum {c in COURSES} easiness[c] = 0;

data;
set STUDENTS := include "names.txt" ;
set COURSES := include "courses.txt" ;
param: GRADES: grade := include "namecoursegrade.txt" ;

solve;
```

Python notebook and three data files are here...

https://vanderbei.princeton.edu/307/python/FairGrading.ipynb

https://vanderbei.princeton.edu/307/models/grades/names.txt

https://vanderbei.princeton.edu/307/models/grades/courses.txt

https://vanderbei.princeton.edu/307/models/grades/namecoursegrade.txt

Student-by-Student Output

```
Alex (4.61)
 2.7   3.9    Statistics (-0.7)
 2.7   3.9    Epistemology (-0.7)
 3.7   4.4    Phrenology (-0.2)
 3.0   4.4    Pharmacology (-0.2)
 3.7   4.3    Philology (-0.4)
 4.3   4.3    Theology (-0.3)
 3.3   3.9    Geology (-0.7)
 4.0   4.0    Gynecology (-0.6)
 3.43  4.13   -- avg -- (-0.48)
---------------------------------
Andy (3.71)
 2.7   4.0    Scientology (0.3)
 2.7   4.0    Etymology (0.3)
 3.7   3.9    Sociology (0.2)
 3.0   4.0    Psychology (0.3)
 3.7   4.2    Cosmology (0.5)
 4.3   4.3    Eulogy (0.6)
 3.3   3.8    Immunology (0.1)
 4.0   4.0    Methodology (0.3)
 3.43  4.03   -- avg -- ( 0.32)
---------------------------------
Ariel (4.03)
 3.0   3.3    Epistemology (-0.7)
 3.3   4.3    Topology (0.2)
 4.3   4.3    Etymology (0.3)
 4.3   4.6    Eulogy (0.6)
 3.7   4.0    Genealogy (-0.1)
 4.3   4.4    Morphology (0.3)
 3.7   4.5    Pathology (0.4)
 4.3   4.4    Technology (0.3)
 3.86  4.21   -- avg -- ( 0.18)
---------------------------------
Billy (4.38)
 4.3   4.6    Etymology (0.3)
 4.3   4.7    Psychology (0.3)
 4.3   4.3    Apology (-0.1)
 4.3   4.9    Cosmology (0.5)
 3.0   4.0    Philology (-0.4)
 4.3   4.5    Ideology (0.1)
 4.3   4.5    Immunology (0.1)
 4.3   4.7    Methodology (0.3)
 4.14  4.52   -- avg -- ( 0.15)
---------------------------------
Bobby (4.08)
 4.3   4.3    Topology (0.2)
 3.3   3.7    Philology (-0.4)
 4.3   4.7    Eulogy (0.6)
 4.0   4.0    Genealogy (-0.1)
 4.3   4.5    Pathology (0.4)
 4.3   4.5    Urology (0.4)
 4.08  4.28   -- avg -- ( 0.21)
---------------------------------
Brett (4.43)
 4.3   4.7    Topology (0.2)
 3.7   4.2    Pharmacology (-0.2)
 4.3   4.7    Demonology (0.3)
 4.3   4.3    Ecology (-0.1)
 4.3   4.5    Ideology (0.1)
 4.0   4.0    Ontology (-0.4)
 4.3   4.8    Technology (0.3)
 4.17  4.45   -- avg -- ( 0.02)
---------------------------------
Brook (3.78)
 4.0   4.0    Topology (0.2)
 2.7   3.6    Phrenology (-0.2)
 4.0   4.3    Cosmology (0.5)
 2.3   3.4    Philology (-0.4)
 3.0   3.5    Theology (-0.3)
 3.7   4.0    Demonology (0.3)
 3.7   4.1    Technology (0.3)
 4.3   4.3    Terminology (0.5)
 3.46  3.91   -- avg -- ( 0.13)
---------------------------------
Cameron (3.73)
 4.0   4.0    Etymology (0.3)
 4.0   4.0    Demonology (0.3)
 4.0   4.3    Eulogy (0.6)
 3.3   3.6    Ecology (-0.1)
 3.0   3.3    Ontology (-0.4)
 4.0   4.2    Pathology (0.4)
 3.3   3.7    Tautology (-0.1)
 3.7   4.1    Technology (0.3)
 3.66  3.89   -- avg -- ( 0.16)
---------------------------------
Cary (4.35)
 4.3   4.6    Psychology (0.3)
 4.0   4.0    Philology (-0.4)
 4.3   4.9    Eulogy (0.6)
 4.3   4.6    Methodology (0.3)
 4.3   4.7    Morphology (0.3)
 4.3   4.3    Tautology (-0.1)
 4.3   4.7    Urology (0.4)
 4.26  4.57   -- avg -- ( 0.21)
---------------------------------
```

Full set of output is here.

Course-by-Course Output

```
Statistics (-0.70)
 2.7   3.9    Alex (4.6)
 3.7   3.7    Casey (4.4)
 3.7   3.7    Dale (4.4)
 3.0   3.8    Darcy (4.5)
 2.7   3.5    Emerson (4.2)
 2.3   3.0    Esme (3.7)
 2.7   3.3    Harley (4.0)
 2.3   2.9    Jordan (3.6)
 4.0   4.0    Max (4.7)
 3.7   3.8    Porntip (4.5)
 2.3   2.9    Sunny (3.6)
 3.3   3.5    Sydney (4.2)
 3.03  3.50   -- avg -- ( 4.20)
-----------------------------------
Epistemology (-0.73)
 2.7   3.9    Alex (4.6)
 3.0   3.3    Ariel (4.0)
 2.7   3.0    Jade (3.7)
 2.0   2.9    Lou (3.7)
 4.0   4.0    Max (4.7)
 2.3   3.3    Robin (4.0)
 2.7   2.7    Sam (3.4)
 2.3   2.8    Tracy (3.6)
 2.71  3.24   -- avg -- ( 3.97)
-----------------------------------
Scientology (0.29)
 2.7   4.0    Andy (3.7)
 4.3   4.7    Casey (4.4)
 4.3   4.7    Dale (4.4)
 4.3   4.4    Dominique (4.1)
 3.7   3.9    Esme (3.7)
 4.3   4.8    Lindsey (4.5)
 4.3   4.4    Meryl (4.1)
 4.3   4.3    Robin (4.0)
 4.3   4.5    Sydney (4.2)
 4.06  4.40   -- avg -- ( 4.11)
-----------------------------------
Astrology (-0.01)
 4.0   4.5    Darcy (4.5)
 4.3   4.7    Daryl (4.7)
 3.7   3.7    Jade (3.7)
 4.3   4.3    Morgan (4.3)
 3.0   3.6    Sunny (3.6)
 3.86  4.16   -- avg -- ( 4.17)
-----------------------------------
Topology (0.22)
 3.3   4.3    Ariel (4.0)
 4.3   4.3    Bobby (4.1)
 4.3   4.7    Brett (4.4)
 4.0   4.0    Brook (3.8)
 4.3   4.5    Chris (4.2)
 4.3   5.0    Drew (4.7)
 3.3   3.8    Jordan (3.6)
 4.0   4.4    Kelly (4.1)
 4.0   4.5    Morgan (4.3)
 4.3   4.3    Reese (4.1)
 4.3   4.9    Skye (4.7)
 4.3   4.4    Sydney (4.2)
 4.06  4.41   -- avg -- ( 4.18)

-----------------------------------
Phrenology (-0.16)
 3.7   4.4    Alex (4.6)
 2.7   3.6    Brook (3.8)
 4.0   4.1    Chris (4.2)
 4.3   4.3    Darcy (4.5)
 3.0   3.6    Devyn (3.8)
 3.3   3.5    Jade (3.7)
 4.3   4.3    Lindsey (4.5)
 3.7   4.6    Max (4.7)
 3.3   3.9    Meryl (4.1)
 3.3   4.1    Morgan (4.3)
 2.0   3.4    Sunny (3.6)
 3.7   3.7    Tanner (3.9)
 3.44  3.97   -- avg -- ( 4.13)
-----------------------------------
Pharmacology (-0.23)
 3.0   4.4    Alex (4.6)
 3.7   4.2    Brett (4.4)
 4.0   4.0    Chris (4.2)
 3.0   3.9    Kelly (4.1)
 3.7   4.0    Kim (4.2)
 4.3   4.3    Porntip (4.5)
 3.3   4.0    Sydney (4.2)
 3.57  4.10   -- avg -- ( 4.33)
-----------------------------------
Etymology (0.27)
 2.7   4.0    Andy (3.7)
 4.3   4.3    Ariel (4.0)
 4.3   4.6    Billy (4.4)
 4.0   4.0    Cameron (3.7)
 3.0   4.1    Dara (3.8)
 3.7   4.3    Harley (4.0)
 3.7   3.9    Lou (3.7)
 4.3   4.3    Peyton (4.0)
```

For full list, click here.

Data Envelopment Analysis

Baseball Batting Statistics

From the web:

Stats on Baseball players who won either an MVP or CyYoung award and was still playing in 2007

	Year	Ag	Tm	Lg	G	AB	R	H	2B	3B	HR	RBI	BA	OBP	SLG	TB
Albert Pujols	2007	27	STL	NL	158	565	99	185	38	1	32	103	.327	.429	.568	321
Dustin Pedroia	2007	23	BOS	AL	139	520	86	165	39	1	8	50	.317	.380	.442	230
Jim Rollins	2007	28	PHI	NL	162	716	139	212	38	20	30	94	.296	.344	.531	380
Alex Rodriguez	2007	31	NYY	AL	158	583	143	183	31	0	54	156	.314	.422	.645	376
Ryan Howard	2007	27	PHI	NL	144	529	94	142	26	0	47	136	.268	.392	.584	309
Justin Morneau	2007	26	MIN	AL	157	590	84	160	31	3	31	111	.271	.343	.492	290
Barry Bonds	2007	42	SFG	NL	126	340	75	94	14	0	28	66	.276	.480	.565	192
Vladimir Guerrero	2007	31	LAA	AL	150	574	89	186	45	1	27	125	.324	.403	.547	314
Miguel Tejada	2007	33	BAL	AL	133	514	72	152	19	1	18	81	.296	.357	.442	227
Ichiro Suzuki	2007	33	SEA	AL	161	678	111	238	22	7	6	68	.351	.396	.431	292
Jeff Kent	2007	39	LAD	NL	136	494	78	149	36	1	20	79	.302	.375	.500	247
Jason Giambi	2007	36	NYY	AL	83	254	31	60	8	0	14	39	.236	.356	.433	110
Chipper Jones	2007	35	ATL	NL	134	513	108	173	42	4	29	102	.337	.425	.604	310
Ivan Rodriguez	2007	35	DET	AL	129	502	50	141	31	3	11	63	.281	.294	.420	211
Sammy Sosa	2007	38	TEX	AL	114	412	53	104	24	1	21	92	.252	.311	.468	193
Ken Griffey	2007	37	CIN	NL	144	528	78	146	24	1	30	93	.277	.372	.496	262

Most numbers are raw data.
But some are derived: BA=H/AB and SLG=TB/AB
Some numbers are *inputs*: "number of games" (G) and "at bats" (AB).
Others are *outputs*: H, R, HR, etc.
See http://en.wikipedia.org/wiki/Baseball_statistics for full list.

How to Combine Stats

Let player i decide how much weight to put on each stat:

$$\text{eff}(i) = \frac{x_{\text{H}}\,\text{H}(i) + x_{\text{R}}\,\text{R}(i) + \cdots + x_{\text{RBI}}\,\text{RBI}(i)}{\text{AB}(i)}.$$

But, don't let the player choose weights that would give him or anyone else a score greater than, say, 100:

$$\text{eff}(j) = \frac{x_{\text{H}}\,\text{H}(j) + x_{\text{R}}\,\text{R}(j) + \cdots + x_{\text{RBI}}\,\text{RBI}(j)}{\text{AB}(j)} \leq 100$$

for all players j in the set of players.

Let the player optimize over all possible choices of weights.

AMPL Model

```
set METRICS;
set PLAYERS;

param AtBats {PLAYERS};          # input (at bats)
param score {PLAYERS, METRICS}; # outputs (by player,type)

var wt {METRICS} >= 0; # weights

param p0 symbolic within PLAYERS; # a specific player

maximize eff: sum {m in METRICS} score[p0,m]*wt[m]/AtBats[p0];

subject to other_eff {p in PLAYERS}:
        sum {m in METRICS} score[p,m]*wt[m]/AtBats[p] <= 100;
```

AMPL Data

```
data;

param: PLAYERS:          AtBats :=
'Albert Pujols'          565
'Dustin Pedroia'         520
'Jim Rollins'            716
'Alex Rodriguez'         583
'Ryan Howard'            529
'Justin Morneau'         590
'Barry Bonds'            340
'Vladimir Guerrero'      574
'Miguel Tejada'          514
'Ichiro Suzuki'          678
'Jeff Kent'              494
'Jason Giambi'           254
'Chipper Jones'          513
'Ivan Rodriguez'         502
'Sammy Sosa'             412
'Ken Griffey'            528
;
```

AMPL Data Continued

```
set METRICS := R    H   2B 3B  HR  RBI  SB  BB  TB;

param score:             R    H   2B 3B  HR  RBI  SB  BB  TB :=
'Albert Pujols'          99  185  38  1  32  103   2  99 321
'Dustin Pedroia'         86  165  39  1   8   50   7  47 230
'Jim Rollins'           139  212  38 20  30   94  41  49 380
'Alex Rodriguez'        143  183  31  0  54  156  24  95 376
'Ryan Howard'            94  142  26  0  47  136   1 107 309
'Justin Morneau'         84  160  31  3  31  111   1  64 290
'Barry Bonds'            75   94  14  0  28   66   5 132 192
'Vladimir Guerrero'      89  186  45  1  27  125   2  71 314
'Miguel Tejada'          72  152  19  1  18   81   2  41 227
'Ichiro Suzuki'         111  238  22  7   6   68  37  49 292
'Jeff Kent'              78  149  36  1  20   79   1  57 247
'Jason Giambi'           31   60   8  0  14   39   1  40 110
'Chipper Jones'         108  173  42  4  29  102   5  82 310
'Ivan Rodriguez'         50  141  31  3  11   63   2   9 211
'Sammy Sosa'             53  104  24  1  21   92   0  34 193
'Ken Griffey'            78  146  24  1  30   93   6  85 262
;
```

AMPL Solver Call and Print Statements

For each player, solve the optimization problem.

In AMPL:

```
for {p in PLAYERS} {
        let p0 := p;
        solve;
        printf "%6.3f %-20s ", eff, p > "player_eff";
        printf {m in METRICS}: "%7.2f ", wt[m] > "player_eff";
        printf "\n" > "player_eff";
        printf "%-20s(%5.2f): \n", p, eff > "dominators";
        printf {pp in PLAYERS: other_eff[pp].body > eff}:
                " %-20s(%5.2f) \n", pp, other_eff[pp].body
                        > "dominators";
}
```

Looping in AMPL is covered in Chapter 13, *Command Scripts*, of the AMPL book. Here's a link:

http://www.ampl.com/BOOK/CHAPTERS/16-script.pdf

Output: Player_eff

		R	H	2B	3B	HR	RBI	SB	BB	TB
98.5	Albert Pujols	0	263	0	0	0	0	0	70	0
95.0	Dustin Pedroia	0	181	427	0	0	0	416	0	0
100.0	Jim Rollins	22	8	245	703	55	18	695	25	28
100.0	Alex Rodriguez	29	5	144	257	228	39	682	36	28
100.0	Ryan Howard	0	1	40	218	119	259	1	93	3
84.0	Justin Morneau	0	178	0	916	0	164	0	0	0
100.0	Barry Bonds	26	8	55	719	76	34	83	158	25
100.0	Vladimir Guerrero	2	5	613	30	14	200	11	5	9
87.0	Miguel Tejada	0	274	0	0	0	38	0	0	0
100.0	Ichiro Suzuki	7	87	4	4	2	7	1149	43	3
89.4	Jeff Kent	0	278	78	0	0	0	0	0	0
74.7	Jason Giambi	0	248	0	0	156	0	0	48	0
100.0	Chipper Jones	39	13	580	262	107	25	65	43	32
82.8	Ivan Rodriguez	0	278	78	0	0	0	0	0	0
91.8	Sammy Sosa	0	0	502	569	0	274	0	0	0
85.1	Ken Griffey	0	248	0	0	156	0	0	48	0

Output: Player_eff Sorted

		R	H	2B	3B	HR	RBI	SB	BB	TB
100.0	Vladimir Guerrero	2	5	613	30	14	200	11	5	9
100.0	Ryan Howard	0	1	40	218	119	259	1	93	3
100.0	Jim Rollins	22	8	245	703	55	18	695	25	28
100.0	Ichiro Suzuki	7	87	4	4	2	7	1149	43	3
100.0	Chipper Jones	39	13	580	262	107	25	65	43	32
100.0	Barry Bonds	26	8	55	719	76	34	83	158	25
100.0	Alex Rodriguez	29	5	144	257	228	39	682	36	28
98.5	Albert Pujols	0	263	0	0	0	0	0	70	0
95.0	Dustin Pedroia	0	181	427	0	0	0	416	0	0
91.8	Sammy Sosa	0	0	502	569	0	274	0	0	0
89.4	Jeff Kent	0	278	78	0	0	0	0	0	0
87.0	Miguel Tejada	0	274	0	0	0	38	0	0	0
85.1	Ken Griffey	0	248	0	0	156	0	0	48	0
84.0	Justin Morneau	0	178	0	916	0	164	0	0	0
82.8	Ivan Rodriguez	0	278	78	0	0	0	0	0	0
74.7	Jason Giambi	0	248	0	0	156	0	0	48	0

Output—Dominators

```
Albert Pujols        (  98.5):
    Barry Bonds          ( 100.0)
    Chipper Jones        ( 100.0)
Dustin Pedroia       (  95.0):
    Jim Rollins          ( 100.0)
    Alex Rodriguez       (  96.6)
    Ichiro Suzuki        ( 100.0)
    Chipper Jones        ( 100.0)
Jim Rollins          ( 100.0):
Alex Rodriguez       ( 100.0):
Ryan Howard          ( 100.0):
Justin Morneau       (  84.0):
    Albert Pujols        (  90.0)
    Jim Rollins          ( 100.0)
    Alex Rodriguez       ( 100.0)
    Ryan Howard          (  90.2)
    Vladimir Guerrero    (  95.2)
    Ichiro Suzuki        (  88.6)
    Chipper Jones        ( 100.0)
    Sammy Sosa           (  84.0)
Barry Bonds          ( 100.0):
Vladimir Guerrero    ( 100.0):
.
.
.
```

Data Envelopment Analysis (DEA)

Baseball example is an illustration of a technique called *Data Envelopment Analysis*.
In general, there can be several inputs and outputs.
Technique has been used widely:

- health care
- education
- banks
- manufacturing
- benchmarking
- management evaluation
- fast food restaurants
- retail stores

Reference:

http://en.wikipedia.org/wiki/Data_Envelopment_Analysis