Modeling in Context: Teaching Confounding and Adjustment through the Common Core Standards

Milo Schield & Daniel T. Kaplan
Augsburg College and Macalester College

Audience and Objective

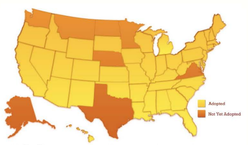
This poster is directed to high-school teachers and curriculum developers. Our objective is to draw attention to new possibilities for teaching statistical literacy concepts through the Common Core Standards, and to show how examples based on confounding and adjustment can provide concrete and compelling settings for developing many of the mathematical, statistical, and modeling components of the Common Core.

The Common Core Standards

COMMON CORE
STATE STANDARDS FOR

Mathematics

***Mission Statement**: "The Common Core State Standards provide a consistent, clear understanding of what students are expected to learn, so teachers and parents know what they need to do to help them. The standards are designed to be **robust and relevant to the real world**, reflecting the knowledge and skills that our young people need for success in college and careers. With American students fully prepared for the future, our communities will be best positioned to compete successfully in the global economy." [emph. added]*



US States that have adopted the Common Core as of May 2011

"Robust and Relevant to the Real World"

Any claim of teaching mathematics that is "robust and relevant to the real world" ought to include some consideration of complexity: that there are generally multiple factors that contribute to an outcome.

In elementary statistics, such complexity is often related to the ideas of "confounding" and "lurking variables." Unfortunately, the analysis statistical methods taught presume that confounding and lurking variables can be ignored.

Instead, we ought to be teaching not only about the perils of confounding, but also about ways of dealing with it.

The linking of mathematics and statistics in the Common Core provides an opportunity to build simple models of real-world situations that illustrate both the influence of confounding and methods for adjusting for it.

Some Examples of Confounding

- The annual death rate in the US (at 9 per 1000) is higher than that in Mexico (5 per 1000). Does this mean that Mexico is a healthier/safer place than the US? The confounder: a different mix of ages. The Mexican population is much younger than the US population.
  Useful health statistics are always "age adjusted" to avoid the misleading impression given by "raw rates."
- US SAT scores improved from 1981 to 2002 every ethnic/racial group: Blacks (+9%), Asians (+27%), Mexicans (+7%), Puerto Ricans (+18%), Whites (+8%), and American Indians (+7%).
  Looking at the overall average scores obscures this improvement; the aggregate scores were flat. The confounder? The changing ethnic/racial demographics of school children.
- 10% of statistically- significant differences between state National Assessment of Education Process (NAEP) scores were reversed after controlling for a confounder. [See Ref. 7.]

Modeling in the Common Core

Mathematics | High School—Modeling

Modeling links classroom mathematics and statistics to everyday life, work, and decision-making. Modeling is the process of choosing and using appropriate mathematics and statistics to analyze empirical situations, to understand them better, and to improve decisions. Quantities and their relationships in physical, economic, public policy, social, and everyday situations can be modeled using mathematical and statistical methods. When making mathematical models,

Note the emphasis on improving decision-making and making connections to the world of economic, policy, and social situations. Learning about confounding and adjustment is an important component of that connection.

Topics from the Common Core Standards

There is the potential for a strong connection between the mathematics part of the Common Core and the Statistics/Probability component. That connection is through **confounding** and **adjustment**, topics that help make math and statistics "robust and relevant to the real world."

Mathematics | Grade 8

In Grade 8, instructional time should focus on three critical areas: (1) formulating and reasoning about expressions and equations, including modeling an association in bivariate data with a linear equation, and solving linear equations and systems of linear equations; (2) grasping the concept of a function and using functions to describe quantitative relationships; (3) analyzing two- and three-dimensional space and figures using distance, angle, similarity, and congruence, and understanding and applying the Pythagorean Theorem.

By Grade 8, students are expected to master the linear description of bivariate patterns. The primary school-level standards refer to breaking data up into groups. Putting these two together is all that's required to develop a substantial understanding of **confounding** and **adjustment**. Alas, the words **confounding**, **adjustment**, or **standardization** never appear in the Common Core standards, even though all the components needed to understand them are there and even though they relate strongly to the Modeling theme of the Common Core.

Mathematics | High School—Statistics and Probability★

Decisions or predictions are often based on data—numbers in context. These decisions or predictions would be easy if the data always sent a clear message, but the message is often obscured by variability. Statistics provides tools for describing variability in data and for making informed decisions that take it into account.

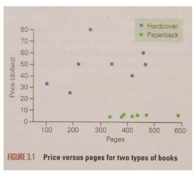
The Common Core statistics components emphasizes variability in terms of distribution and spread. This is, of course, central to statistics. But the emphasis on traditional introductory statistics topics misses the opportunity to build on the Common Core mathematics components in a way that will enhance their robustness and relevance to the real world. For example:

Connections to Functions and Modeling. Functions may be used to describe data; if the data suggest a linear relationship, the relationship can be modeled with a regression line, and its strength and direction can be expressed through a correlation coefficient.

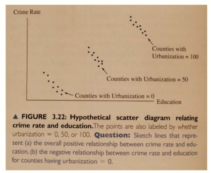
The correlation coefficient is a traditional and important topic, but as the examples below show, it can be misleading on its own.

Statistics Textbooks and Confounding

The terms "confounding" and "lurking variable" are commonly found in introductory statistics textbooks, which often contain examples of how a correlation can give a misleading picture. Here are two examples:



Utts & Heckard, p. 59



Agresti & Franklin, p. 130

But although its an important topic, even college-level statistics textbooks rarely engage confounding beyond the statement, "Beware!"

The following table shows the number of index entries in each of 6 major introductory statistics texts (identified as A-F) for each of several words relating to confounding and causation:

Index Term	Confounding	Lurking Variable	Case-Control	Control Variable	**Adjustment**	**Standardization**	**Covariate**	Causation	Simpson's Paradox
A	13	2	4	0	0	0	0	14	2
B	6	10	1	0	0	0	0	13	2
C	5	12	5	1	0	0	0	6	4
D	4	4	0	0	0	0	0	4	0
E	4	3	0	0	0	0	0	1	2
F	0	0	0	0	0	0	0	6	0

The importance of giving students the tools they need to analyze confounding situations is well described by Nicholson, Ridgway, and McCusker [ref. 6]:

> *"Working with visual representations of multivariate data at an early stage would help students to develop mental models of possible relationships between multiple variables which would give them a stronger conceptual basis for considering the formal statistical analysis they will meet in courses such as psychology or geography."*

A Linear Modeling Approach

Imagine comparing two baseball players, Ted and Sam.

Their batting averages:

Ted	0.340
Sam	0.260

Who is the better batter? Ted, obviously.

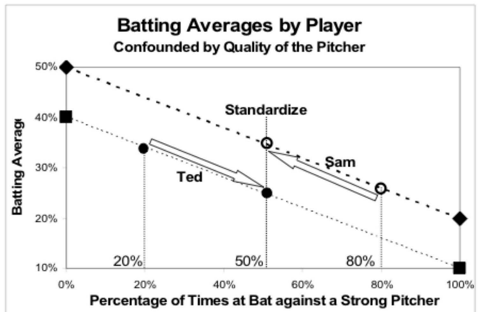
But any school child interested in baseball will know enough to see that it depends. How well a batter does depends not just on the batter, but on the pitcher and on the defensive strength of the team.

The quality of the pitcher makes a big difference; that's why pitchers are so important to baseball teams. It seems obvious to compare the two players under the same circumstances, leading to a stratified table, like this:

	Weak Pitcher	Strong Pitcher
Ted	0.400	0.100
Sam	0.500	0.200

Against either a strong pitcher or a weak one, Sam is the stronger player. But not overall.

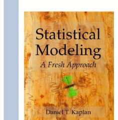
This is Simpson's Paradox. But it's not really a paradox at all. It's easy to understand why Ted appears better than Sam in the overall records. This can be explained intuitively, but also quantitatively using the mathematical concepts contained in the Common Core.



- Make a graph of the batting average as a function of how often the batter faces a strong pitcher.
- We know that when Sam faces a strong pitcher, his batting average is 0.200. Against a weak pitcher, it's 0.500. Plot out those two bits of data — the diamond-shaped points.
- Over the season, Sam faces a variety of pitchers, some strong and some weak. His overall batting average is therefore a mixture of his performance against the strong and weak pitchers. That's the heavy dashed line. As it happens, Sam played in a league with good pitching: he faced strong pitchers 80% of the time. Thus his batting average was between his performance for strong and weak pitchers, but closer to that for strong pitchers.
- A similar analysis is done for Ted, who plays in a league with few strong pitchers.
- The two lines are models describing how each player will perform against different mixtures of pitchers.
- It's evident from the lines themselves that Sam is the better batter.
- Standardization is the process of using the models to compare Sam and Ted on an equal basis, say a 50% mixture of strong and weak pitchers.

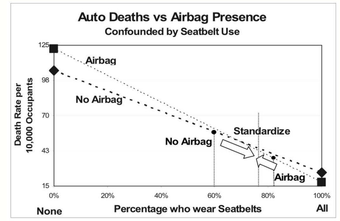
The modeling process lets us take into account the confounding variable of pitcher strength. Standardization lets us compare the batters on an equal footing.

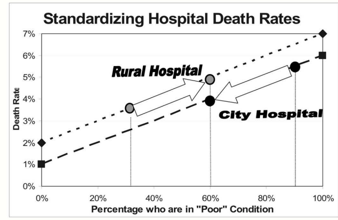
Toward a University-Level Understanding

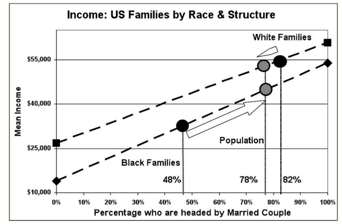


The approach to confounding and adjustment presented here builds on skills already existing in the Common Core. With college-level mathematical tools, a more general approach can be taken, allowing for the consideration of multiple variables. Such modeling approaches are more consistent with professional practice and the need to adjust for confounding in addressing real-world issues. One such text is *Statistical Modeling: A Fresh Approach*.

Other Settings for Modeling and Adjustment







References

1. The Common Core State Standards for Mathematics. (2010) `corestandards.org`
2. Peck, R. (2009). *Making Sense of Statistical Studies*
3. Lesser, L. (2001). *Representations of Reversal: Exploring Simpson's Paradox.* In Albert A. Cuoco and Frances R. Curcio (Eds.). The Roles of Representation in School Mathematics, pp. 129-145 [chapter in 2001 NCTM yearbook].
4. Gould, R. (2011). *Eat Less Salt, Drink More Wine, ... Teaching Students to Understand the Role Of Data Collection in Statistical Inference.* CAUSE Web Webinar: Teaching. And Learning. April 12, 2011. `www.causeweb.org/webinar/teaching`
5. Schield, M. (2006). Presenting Confounding and Standardization Graphically. STATS Magazine, American Statistical Association. Fall 2006. pp. 14-18. `www.StatLit.org/pdf/2006SchieldSTATS.pdf`
6. Nicholson, Ridgway and McCusker (2009). *Luring non-quantitative majors into advanced statistical thinking (and luring statistical educators into real statistics)* . ICOTS-8 `www.statlit.org/pdf/2010NicholsonRidgwayMcCuskerICOTS.pdf`
7. Terwilliger, J. and Schield, M. "Frequency of Simpson's Paradox in NAEP Data"

Contacts:

Milo Schield
`schield@augsburg.edu`

Daniel Kaplan
`kaplan@macalester.edu`

See also: `www.StatLit.org` & `www.mosaic-web.org`

The authors gratefully acknowledge the long-term support of the W.M. Keck Foundation for the *Statistical Literacy* project at Augsburg College and the *Data Fluency* initiative at Macalester College.

Kaplan: *Dept. of Mathematics, Statistics, and Computer Science, Macalester College* Schield: *Dept. of Business Administration, Augsburg College* Mail: `kaplan@macalester.edu` & `schield@augsburg.edu` WWW: `www.mosaic-web.org` and `www.StatLit.org`