2017 IARIW-BOK

The Dimensions of Ordinal Well-Being Indexes: Using Orthogonal Weighting with the Kids Count Index

April 27, 2017

Session 5A: Construction of Indexes of Well-Being

Ross Knippenberg
Assistant Director of Economics
American Veterinary Medical Association
Chicago, IL, USA

AVMA™
Our Passion. Our Profession.

I: Introduction

Ross Knippenberg
American Veterinary Medical Association

America's Best States to Live In

Thomas C. Frohlich, Michael B. Sauter and Samuel Stebbins, 24/7 ...

Americans often evaluate their home states according to such subjective conditions as climate preference, the presence of friends and family, and personal history. In addition to these subjective measures, more objective socioeconomic factors also contribute to life satisfaction. It is such quantitative measures that can help assess the broader quality of life in a given state.

24/7 Wall St. reviewed three statewide social and economic measures -- poverty rate, educational attainment, and life expectancy at birth -- to rank each state's living conditions. Socioeconomic outcomes vary greatly between states.

Massachusetts, home to one of the nation's wealthiest and most highly educated populations, leads the nation in quality of life. Mississippi, the poorest state in the country, trails the other 49 states.

ALSO READ: Cities Where You Don't Want to Get Sick

While satisfactory living conditions are possible with low incomes, this is true only to a point. Once incomes fall below the poverty line, for example, financial constraints are far more likely to diminish quality of life. New Mexico and Mississippi report poverty rates of over 20%.

https://www.yahoo.com/finance/news/america-best-states-live-110053602.html

AVMA
Our Passion. Our Profession.

Education levels are another major contributor to a community's living conditions -- not just as a basis of economic prosperity, but also as a component of an individual's quality of life. Due in part to the greater access to high paying jobs that often require a college degree, incomes also tend to be higher in these states. In all of the 15 best states in which to live, the typical household earns more than the national median household income of $55,775.

Many of these strong socioeconomic measures lead to better living conditions, which in turn help lead healthier and longer lives -- also used in ranking states. The difference in life expectancy between Mississippi, where people tend to live the shortest lives, and Massachusetts, is 5.6 years. The likelihood of living a relatively long life as a resident of a particular state is closely associated with that state's living conditions.

Housing markets are also indicative of quality of living. A high median home value, for instance, frequently means high demand for housing in the area. Nationwide, the typical home is worth $194,500. In most of the 25 top states, the median home value far exceeds the nationwide median.

Lower home values are indicative of and contribute to relatively affordable costs of living. Of course, low home values are also a product of a lack of demand in a housing market, which is often driven by poor living conditions. The average cost of goods and services in most of the best states to live is greater than the national average, while the average cost of living in all of the 25 states on the lower end of our list is less than the national average.

AVMA
Our Passion. Our Profession.

To identify the best and worst states in which to live, 24/7 Wall St. devised an index composed of three socioeconomic measures for each state: poverty rate, the percentage of adults who have at least a bachelor's degree, and life expectancy at birth. The selection of these three measures was inspired by the United Nations' Human Development Index. Poverty rates and bachelor attainment rates came from the U.S. Census Bureau's 2015 American Community Survey. Life expectancies at birth are from the Centers for Disease Control and Prevention and are as of 2012, latest year for which data is available. Unemployment rates are from the Bureau of Labor Statistics, and are for October 2016, the most recent available month of data.

These are the best and worst states to live in.

50. Mississippi

1 > **10-yr. population growth:** 6.0% (11th lowest)
2 > **Oct. unemployment rate:** 5.9% (5th highest)
3 > **Poverty rate:** 22.0% (the highest)
4 > **Life expectancy at birth:** 74.5 years (the lowest)

Based on a range of social and economic factors, Mississippi is the worst state to live in. With the nation's highest poverty rate of 22.0% and the lowest life expectancy of 74.5 years, economic factors have likely had an adverse effect on the quality of life of Mississippians. While health insurance coverage has increased dramatically across the nation in recent years, many Americans, especially those in low income families, remain uncovered. The typical household in Mississippi earns $40,593 annually, the lowest of all states and in stark contrast to the national annual household income of $55,775. The percentage of people without health insurance in Mississippi, at 12.7%, is sixth highest of all states.

MA
Profession.

Methodology Comments

1. Only use three indicators, but results list four, and discussion talks about three more
   (Why? Are they arbitrary? Is causation implied?)
2. Indicator selection "…inspired by the UN HDI"
3. Methodology unknown
   1. Method of weighting across indicators not discussed
   2. Are ranks independent across indicators?
   3. Is a one-unit change in rank equal across or even within indicators?

   Like using an ordered logit model when should be an interval regression

AVMA
Our Passion. Our Profession.

Other Examples

- Gallup Healthways Well-Being Index
- Best in Show / Top Dog
- Index of Economic Freedom
- Colorado Innovation Index
- Used Car Trade-In Quality Index
- Index of Globalization
- UN Human Development Index

AVMA
Our Passion. Our Profession.

AVMA
Our Passion. Our Profession.

Table 1: Kids Count Indicators

Economic Well-Being Indicators	1. Children in poverty 2. Children whose parents lack secure employment 3. Children living in households with a high housing cost burden 4. Teens not in school and not working
Education	5. Children not attending preschool 6. Fourth graders not proficient in reading 7. Eighth graders not proficient in math 8. High school students not graduating on time
Health	9. Low-birthweight babies 10. Children without health insurance 11. Child and teen deaths per 100,000 12. Teens who abuse drugs or alcohol
Family and Community	13. Children in single-parent families 14. Children in families where the household head lacks a high school diploma 15. Children living in high-poverty areas 16. Teen births per 1,000

The Kids Count Index

- Methodology
  - 16 indicators collected for all 50 states
    - 50x16 data matrix
  - Each indicator is converted to a z-score (subtract national mean, divide by std dev)
  - z-scores are summed for each state
  - States are ranked from lowest sum of z-scores to highest.
  - The 16 indicators are all negative (higher=bad), so lowest score is best.

The Kids Count Index

- **Consider this**: how alike are:
  - "The percent of children living in poverty" and:
  - "The percent of children living in high poverty areas"?
- Probably high
- What is the latent variable?
- Methodology does not account for this

AVMA
Our Passion. Our Profession.

The Kids Count Index

- Ideal computation:
  - Household level
  - Complete data for each household
  - An index, not a regression,
    - use PCA to find primary axes
  - Reweight vars along PCs
  - Reweight survey obs to represent state population
  - Sum up vars as before
- No such dataset exists

AVMA Our Passion. Our Profession.

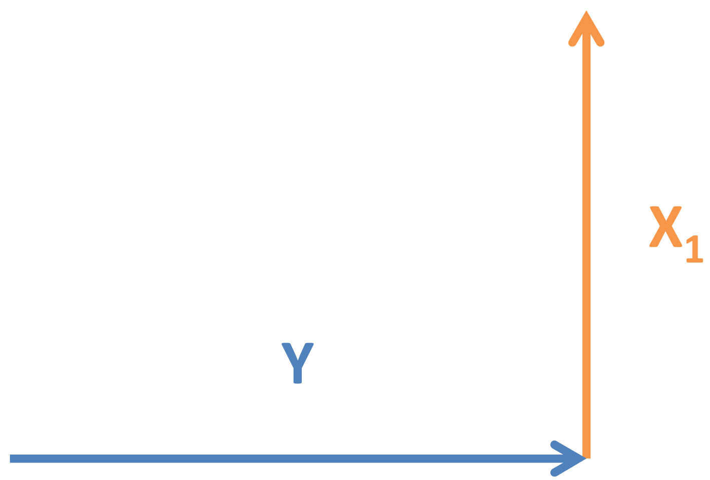
II: Methodology

Ross Knippenberg
American Veterinary Medical Association

AVMA Our Passion. Our Profession.

Distance Metric Concepts

- When adding z-scores, index is implicitly using the manhattan distance metric:

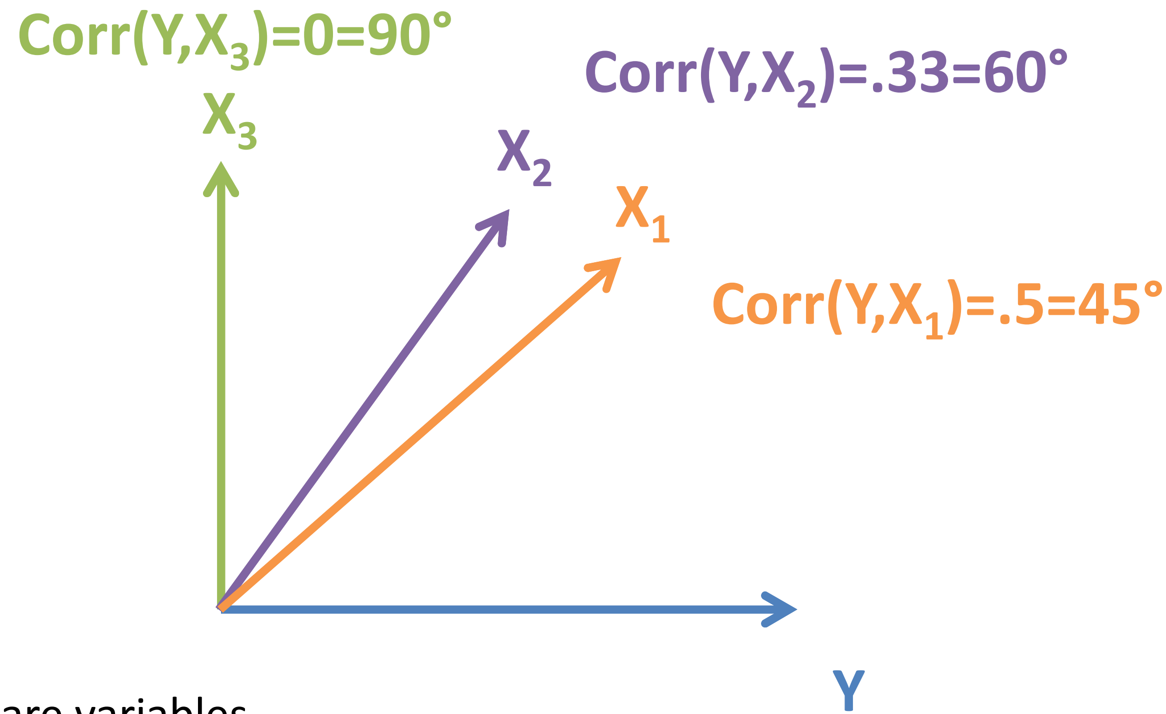


Manhattan Distance = $d(Y, X_1) = Y + X_1$
$= 1 + 1 = 2$

AVMA Our Passion. Our Profession.

Distance Metric Concepts

- Correlation describes the orientation of vectors



Facts:

1.) Y, X_1, X_2 and X_3 are variables

2.) Y is orthogonal to X_3, but not to X_1 or X_2

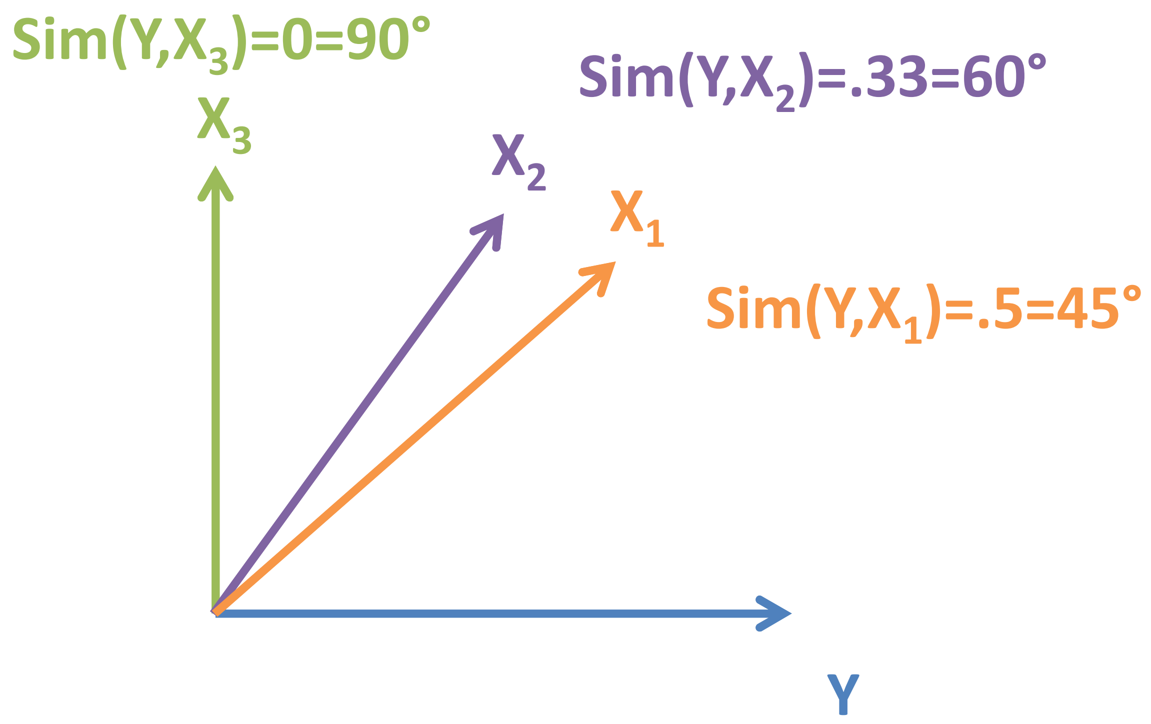
3.) Y and X_3 are separate dimensions

4.) X_1 and X_2 are redundant, linear combinations of Y, and X_3

AVMA Our Passion. Our Profession.

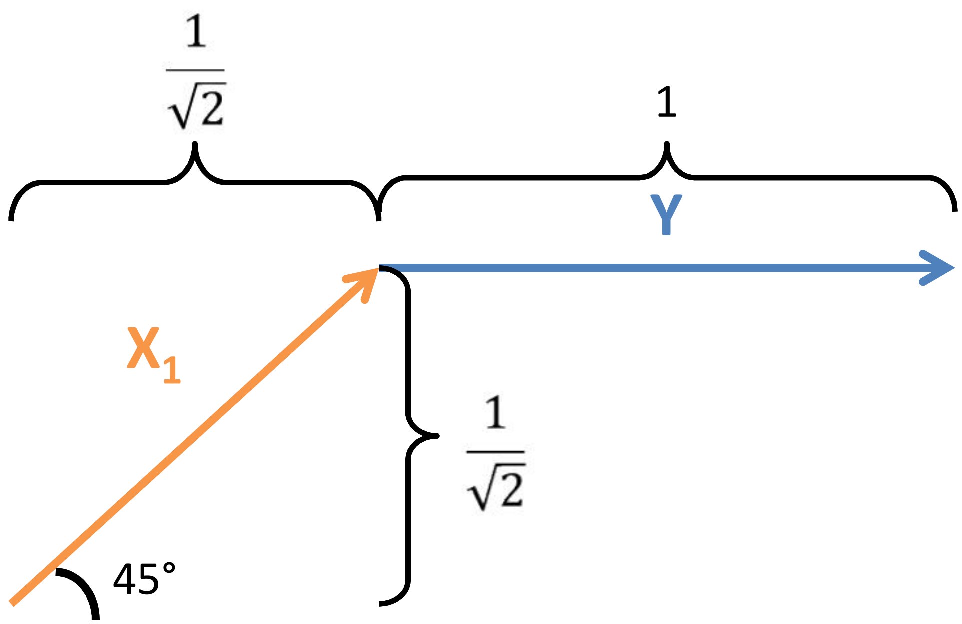
Distance Metric Concepts

- Likewise, similarity describes vector orientation



Distance Metric Concepts

- If Y and X_1 are **not** orthogonal, then **NO** standard distance metric can be correct:



Manhattan Distance ≠ Y + X_1

$$= 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \approx 2.414$$

$$\neq 2$$

AVMA Our Passion. Our Profession.

Motivation

- My basic argument:
  - Dimension ≠ Variable
  - Why?
    - By definition, dimensions are orthogonal
    - Variables may have pairwise correlation
    - Nonzero correlation means non-orthogonal
  - A variable may be it’s own dimension, but not necessarily. Too strong to assume.
  - Distance metrics are based on dimensions **NOT** variables

Motivation

- So how do we reconcile?
  With an orthogonalization technique:
  - Regression Analysis
  - Principal Components Analysis
  - Factor Analysis
  - Multiple Correspondence Analysis
  - Principal Coordinates Analysis
  - Multidimensional Scaling
  - Linear Discriminant Analysis
  - Canonical Correlation Analysis

AVMA
Our Passion. Our Profession.

Motivation

- However, these work only in some situations
- ie., when correlation is non-spurious
  - But how to orthogonalize when using compositional data? (e.g. expenditure data)
  - Compositional data has spurious correlation because always positive.
  - New methodology

Distance Metric Concepts: Euclidean Distance

2-dimensional case:

For orthogonal (perpendicular) Y and X_1:

$$\rho^2 = Y^2 + X_1^2$$

For non-orthogonal Y and X_1 (Law of Cosines):

$$\rho^2 = Y^2 + X_1^2 + 2YX_1\cos(\left(1 - \phi_{Y,X1}\right)90\frac{\pi}{180}))$$

n-dimensional case:

$$\rho^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j \cos(\left(1 - \phi_{i,j}\right)90\frac{\pi}{180}))$$

$$\rho^2 = \vec{\boldsymbol{X}}^T \cos((\boldsymbol{1} - \boldsymbol{\Phi})90\frac{\pi}{180})\,\vec{\boldsymbol{X}}$$

AVMA
Our Passion. Our Profession.

Methodology

1. Estimate similarity between the 16 indicators to estimate network of well-being
2. Convert similarity to distance in degrees (or radians) and arrange into matrix, Φ
3. Convert indicators to z-scores
4. Find distance from national average (the zero vector, $\vec{0}$) to each state, using the Law of Cosines to calculate Euclidean distance:

$$y_i = (\vec{x_i})^T \cos(\Phi)\, \vec{x_i}$$

AVMA
Our Passion. Our Profession.

Methodology

To estimate the angle (similarity) between variables, $\phi_{i,j}$, can use correlation, but I follow a network approach:

1. Calculate which indicators in which states are above the national average
2. For all those states with indicator *i* above average, what percent of *j* are above average? Then vice versa (*j* given *i*). Then find the minimum of the two.

AVMA
Our Passion. Our Profession.

2017 IARIW-BOK

III: Results

Ross Knippenberg
American Veterinary Medical Association

AVMA
Our Passion. Our Profession.

Similarity Matrix

Table 2: Similarity Matrix Φ

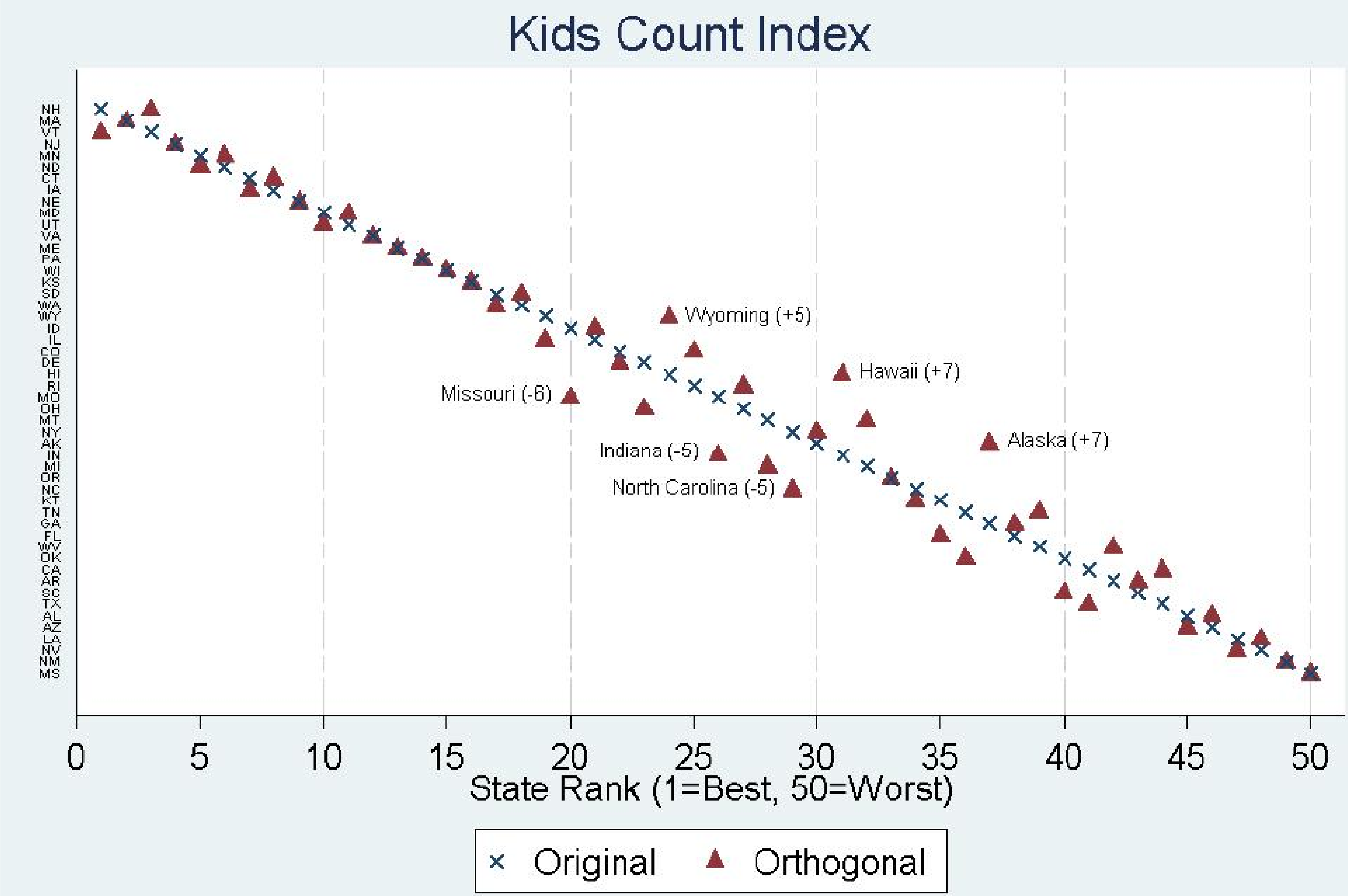
12. % teens who abuse drugs/alcohol

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	0.75	0.46	0.54	0.58	0.68	0.74	0.71	0.69	0.38	0.46	0.42	0.46	0.73	0.67	0.67
2	0.75	1	0.56	0.47	0.5	0.69	0.66	0.78	0.69	0.38	0.44	0.34	0.56	0.83	0.72	0.59
3	0.46	0.56	1	0.48	0.37	0.52	0.56	0.68	0.62	0.41	0.24	0.38	0.55	0.63	0.56	0.46
4	0.54	0.47	0.48	1	0.6	0.76	0.74	0.75	0.66	0.4	0.48	0.44	0.48	0.67	0.64	0.64
5	0.58	0.5	0.37	0.6	1	0.6	0.44	0.43	0.34	0.37	0.53	0.42	0.42	0.43	0.4	0.54
6	0.68	0.69	0.52	0.76	0.6	1	0.74	0.71	0.55	0.4	0.48	0.44	0.48	0.6	0.64	0.64
7	0.74	0.66	0.56	0.74	0.44	0.74	1	0.79	0.72	0.41	0.52	0.41	0.52	0.73	0.67	0.63
8	0.71	0.78	0.68	0.75	0.43	0.71	0.79	1	0.72	0.43	0.5	0.39	0.54	0.77	0.68	0.61
9	0.69	0.69	0.62	0.66	0.34	0.55	0.72	0.72	1	0.41	0.48	0.38	0.55	0.77	0.69	0.62
10	0.38	0.38	0.41	0.4	0.37	0.4	0.41	0.43	0.41	1	0.48	0.43	0.43	0.57	0.4	0.63
11	0.46	0.44	0.24	0.48	0.53	0.48	0.52	0.5	0.48	0.48	1	0.33	0.35	0.43	0.52	0.63
12	0.42	0.34	0.38	0.44	0.42	0.44	0.41	0.39	0.38	0.43	0.33	1	0.38	0.47	0.36	0.42
13	0.46	0.56	0.55	0.48	0.42	0.48	0.52	0.54	0.55	0.43	0.35	0.38	1	0.8	0.69	0.62
14	0.73	0.83	0.63	0.67	0.43	0.6	0.73	0.77	0.77	0.57	0.43	0.47	0.8	1	0.7	0.6
15	0.67	0.72	0.56	0.64	0.4	0.64	0.67	0.68	0.69	0.4	0.52	0.36	0.69	0.7	1	0.64
16	0.67	0.59	0.46	0.64	0.54	0.64	0.63	0.61	0.62	0.63	0.63	0.42	0.62	0.6	0.64	1
sum	9.93	9.96	8.48	9.75	7.98	9.93	10.27	10.49	9.9	7.51	7.86	7	8.83	10.73	9.97	9.93

2. HH Head lacks secure employment

14. HH Head lacks H.S. diploma

11. Child & teen deaths per 100,000



Source: Annie E. Casey Foundation and author's calculations

Discussion

- Takes correlation or co-occurrence into account
- Can be used on compositional data
  - (PCA must be transformed by log-ratio)
- No problem dealing with zeros
- Equally weights dimensions, not variables
- No need for dimension reduction
- Can use high-level indicators; does not require household-level data
- The higher the number of dimensions, the larger the change in final results

AVMA
Our Passion. Our Profession.

Takeaways

- Variables ≠ Dimensions
- The correlation of variables matters for weighting in an index
- In ordinal indices, a small computation change can make a big difference b/c magnitudes are not even between ranks
- Impacts policy decisions
- Index construction methods impacts non-academic practitioners

AVMA
Our Passion. Our Profession.

2017 IARIW-BOK

Thank you!

Ross Knippenberg
American Veterinary Medical Association

AVMA™ Our Passion. Our Profession.