

A set of graphical models to explore the effects of globalization and development on the prevalence of obesity

Shannon Cebron

Johns Hopkins University

scebron@cis.jhu.edu

November 27, 2012

Summary

Project overview

- Under-explored phenomenon
- Graphical models to represent global connections
- Statistical models to analyze graphical features

The RAND Corporation

- Nonpartisan public policy research organization
- A focus area is health and health care
- Has previously worked with the US Congress

What is obesity?

- Defined as a BMI greater than 30 (kilograms of body weight per meters of height squared)
- Indicator for heart disease, diabetes, and more
- Causes bone density and joint problems and difficulty with everyday tasks

Costs to society

- Leading cause of preventable death
- Drives up health care costs
- Lost hours affect the economy (117 billion USD per year in the USA)

Relationship to globalization

- Increasing at highest rates in developed countries
- Increasing at higher rates in neighbors of developed countries
- Burger King in Accra, Ghana
- Exploding urbanization in places like Burundi

Regression model

- A model to predict current obesity rate based on known data

Interactive experience

- An R package to encapsulate obesity prediction models
- Allows user manipulation of data to predict effects of policy changes
- Includes sample data

Project outline

Data acquisition

- Obesity rate
- Development status
- Bordering nations
- Economic data
- Trading data

Source: CIA World Factbook

Graph construction

- Nations as vertices
- Varying adjacency and edge length definitions for each graph
- Example: Border relationships as adjacencies and distance between capital cities as edge lengths

Environment: Python

Project outline

Data extraction

- Neighborhood sizes
- Shortest path to a developed country

Data analysis

Independent variables (examples)

- GDP per capita
- Education expenditures as a percentage of GDP
- Total internet users
- Number of food import partners
- Degree of separation from a highly developed nation
- Rate of urbanization

Dependent variable: Obesity rate

Data analysis

- R has great capability for regression diagnostics
- Will consider not only linear, but also exponential, polynomial, and other regressions
- An acceptable model will have a significant F-statistic and significant T-tests for slope coefficients
- Model must also meet assumptions of regression

Preliminary work: data analysis

Visualization of these data in RStudio

Preliminary work: data analysis

Simple regression of obesity rate on education expenditures

Preliminary work: data analysis

Plot of obesity rate against education expenditures, overlaid with regression line

Preliminary work: data analysis

Multiple linear regression of obesity rate against all basic independent variables

Preliminary work: data analysis

Scatterplot matrix showing relationship between obesity rate and other independent variables

Discovering nonlinear relationships

- Per person imports has a weak linear relationship with obesity rate ($p=0.269$)
- Log per person imports has a strong linear relationship with obesity rate ($p=0.00000027$)

Preliminary work: graph construction

Graph constructed in Python showing national border relationships

Looking forward

- Future research might including a more in-depth exploration of graphical variables
- Efforts may be taken to find obesity data for more than 190 countries
- More lifestyle data and cultural factors could be considered, such as popularity of sports

The End