Community Health Estimator

A self-service population health estimator

Developed by

Jason Summer

Partners

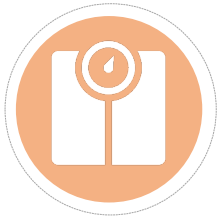
CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

PLACES
LOCAL DATA FOR BETTER HEALTH

With a mission of improving population healthcare at the local level, the app provides a self-service tool enabling community and healthcare officials to better understand population risk and key drivers of community healthcare.

Motivation

Population health is estimated based on 4 community-level categories



Health Outcomes

Health Risk Behaviors

Preventative Measurements

Demographics

Demo

Data Overview

Population health is estimated based on 4 community-level categories

Data Landscape

PLACES is a CDC-led effort that provides model-based, population level community estimates of health measures to US counties.

The app uses a subset of these estimates to predict the proportion of county residents that rate their health as fair/poor.

Health Outcomes:

- Arthritis
- Cancer (except skin)
- Chronic kidney disease
- COPD
- Coronary heart disease
- Asthma
- Depression
- Diabetes
- High blood pressure
- High cholesterol
- Obesity
- Stroke

Risk Behaviors:

- Binge drinking
- Smoking

Preventative Measures:

- Cholesterol screening
- Taking BP medication
- Health insurance
- Annual checkup
- Colorectal cancer screening

Demographics:

- County population
- US region

Relational Database

RDS captures data at 4 points in the overarching pipeline:

1. **Measurement definitions** are captured following database initialization.
2. **Min and max values of scaled parameters**, such as population, are recorded during featurization.
3. **Model coefficients** are captured during training.
4. **User input** is recorded once submitted from the live app, along with their corresponding prediction.

Model Overview

Algorithm

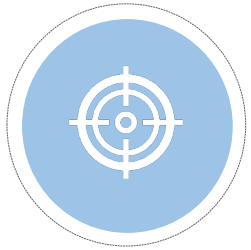
The proportion of county residents
with fair/poor health was treated as a response.

To enable unconstrained linear regression on a
constrained proportion,
the target response was transformed into a log-odds.

Model Success

Model success was based on prediction root mean squared error (RMSE). The model resulted with a RMSE of 0.014, which is superior to the original threshold set at 0.025.

Model Predictions



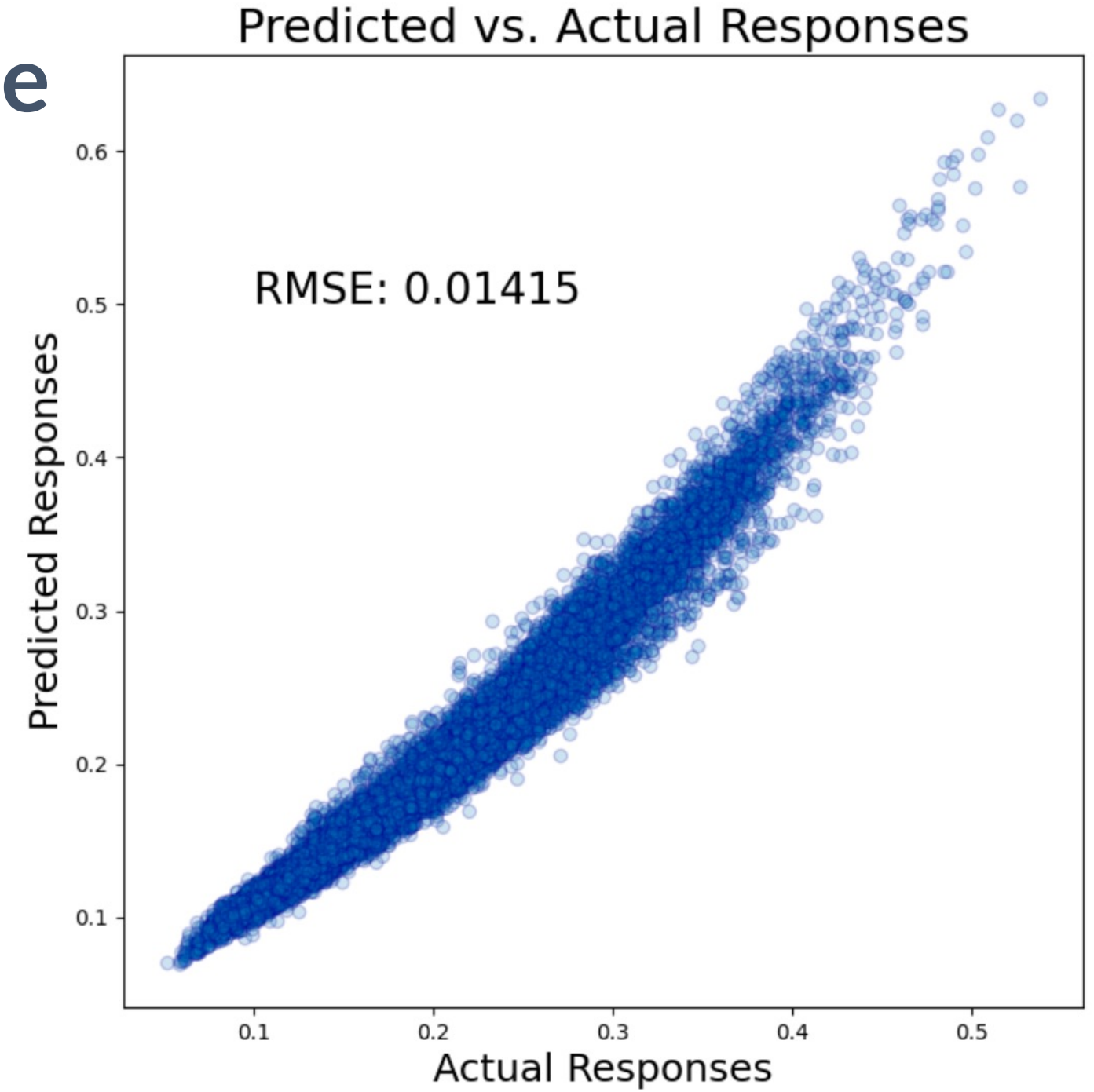
The model serves online predictions based on user input in the online application. Predictions are rendered once the user selects `Estimate`.

Model Performance

There were some measures with significant **multi-collinearity that were removed**.

Other expected correlation remained, e.g., between diabetes and obesity.

However, the array of community estimates remained provided a very effective prediction with a simple model. Predictions had an **average error of 1.4% percentage points** in predicting overall fair/poor health of a community.



Thank You

Jason Summer

jasummer92@gmail.com