

Community Health Estimator

A self-service population health estimator



Developed by

Jason Summer

Partners

 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:

- Population
- US Region

Relational Database

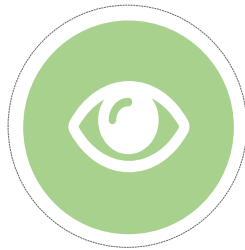
RDS captures data at 4 points in the overarching pipeline:

1. Measurement definitions are captured following database initialization
2. Min and max values of scaled parameters, such as population during featurization
3. Model coefficients during training
4. User input entered in the live app

Model Overview

Algorithm

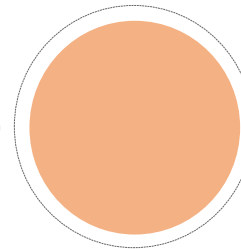
Model success was based on prediction root mean squared error (RMSE). The model resulted with a RMSE of 0.014, which is superior to the original threshold of 0.025 set.



Algorithm

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

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



Model Predictions

The model serves online predictions based on user input in the online application.

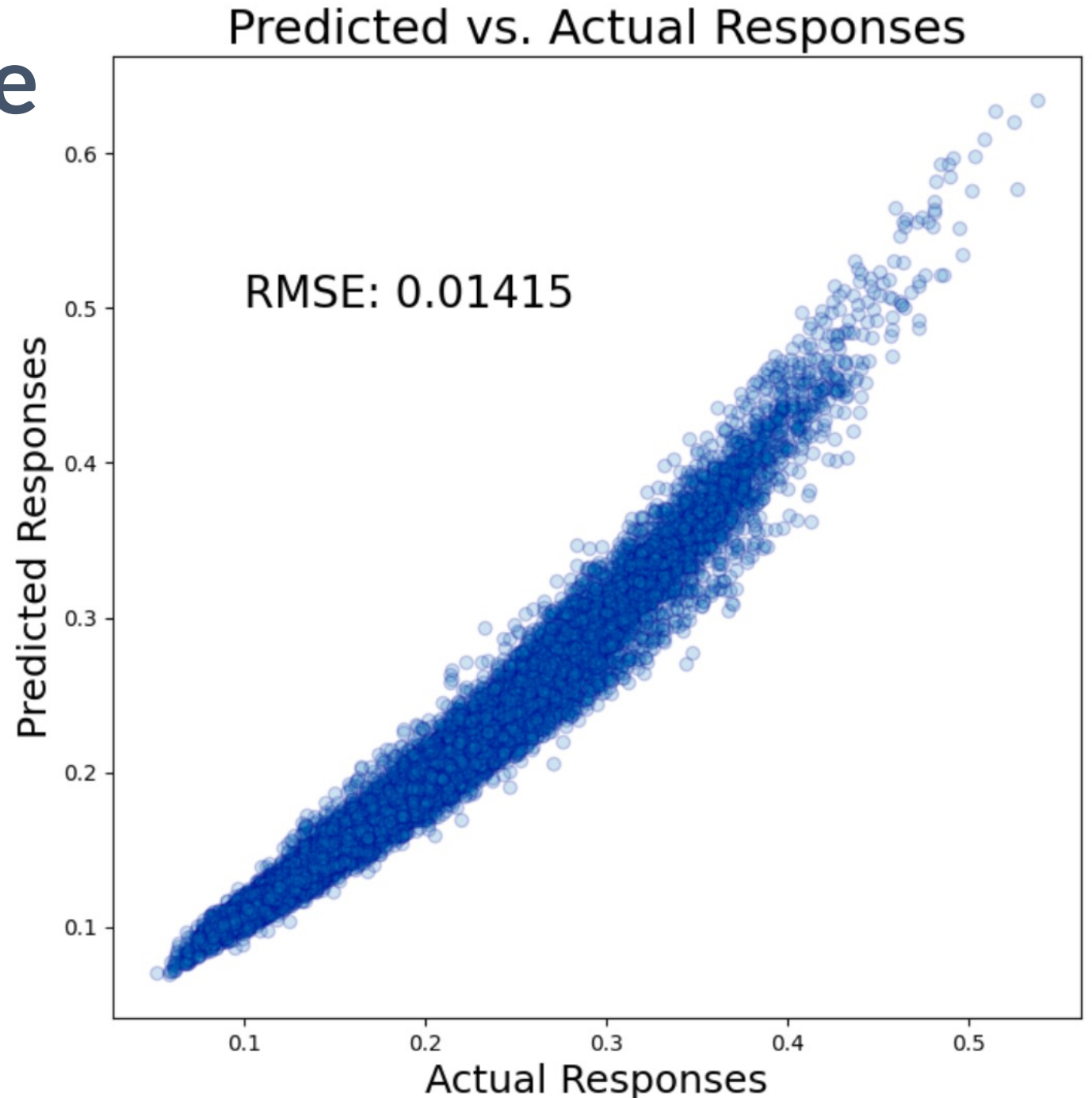


Model Performance

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

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

However, **despite this innate multi-collinearity**, the simple linear model performed very well with an **average error of 1.4% percentage points** in predicting overall fair/poor health of a community.



Thank You



Jason Summer



jasummer92@gmail.com