Author: Rebeca Ansar
Date: September 17, 2022

Executive Summary: Modifying Covid Models 1 & 5

Introduction

Two linear regression models, model 1 and model 5 from “Health in America: What explains the variation in COVID-19 mortality rate across the U.S.?” were modified by changing their dependent variable in the following way: instead of the models predicting cumulative deaths per 1 million county population as of January 23, 2021, the modified models predict the same ratio as of August 15, 2022. No changes were made to the predictors. The two modified models are referred to as ‘modified model 1’ and ‘modified model 5’.

Findings

Both modified models’ adjusted R-squared values increased. Model 1’s adjusted R-squared increased from 0.35 to 0.46 and Model 5’s adjusted R-squared increased from 0.48 to 0.61.

In both the original and modified model 1, there are age, race, socioeconomic, and health features that are statistically significant. All of the same age levels are statistically significant in both the original and modified model 1’s, but ‘a6574’ loses its degree of statistical significance in modified model 1. Only the ‘latino’ race designation is statistically significant in modified model 1. In the original model 1, this race designation along with a few others are statistically significant. Among health factors, ‘disable’, ‘hosp’, ‘vcrime’, and ‘p_liquor’ lose their statistical significance when the original model 1 is modified, i.e. these features are not statistically significant in modified model 1.

In both the original and modified model 5, there are age, race, socioeconomic, and health features that are statistically significant. The most notable difference is found in the ‘statename’ feature. There is no statistically significant state in the original model 5, but 3 states, namely Florida, Nebraska, and Utah are statistically significant in modified model 5. Each of these state levels has a negative coefficient, so keeping all other explanatory variable values static, the predicted cumulative deaths per 1 million county population ratio is lowered if the county belongs to one of these states. Revisiting the age and race categories of features, ‘a5564’ and ‘black’ lose their statistical significance in modified model 5. Within the socioeconomic and health feature categories, ‘mincome’, ‘disable’, ‘drinking’, and ‘lowbirthw’ gain statistical significance in modified model 5. Overall, there are more statistically significant features in modified model 5 than the original model 5.

Conclusion

For both models 1 and 5, the increase in adjusted R-squared indicates that the models are robust, and that many of the explanatory variables included in the original models retained importance and perhaps became even more important as more data on covid deaths was collected over time.

Author: Rebeca Ansar
Date: September 17, 2022

Appendix

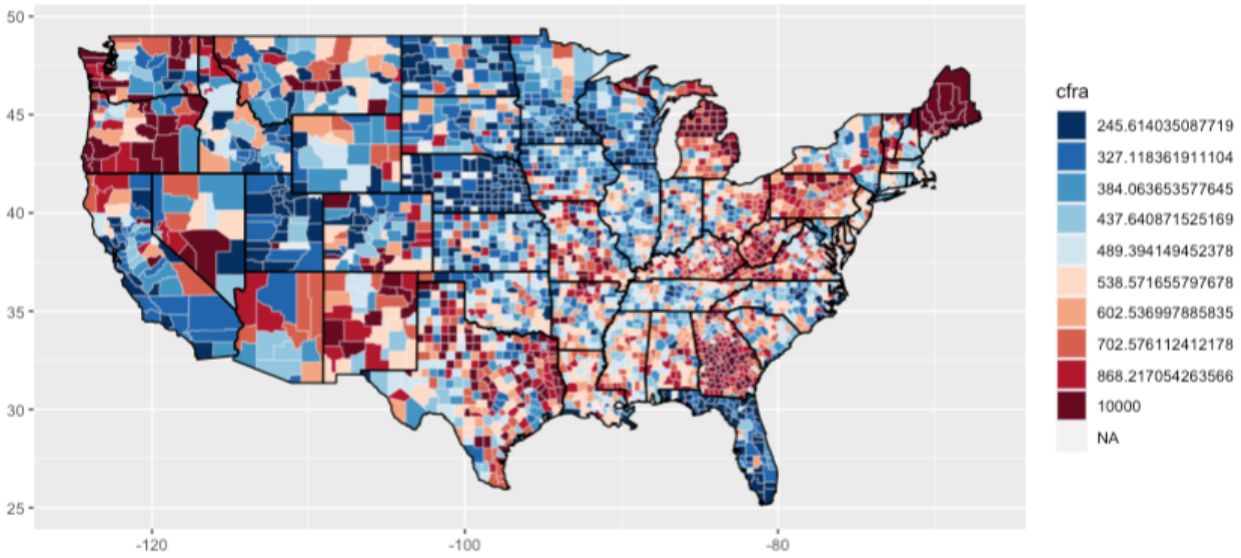
Features whose statistical significance changed between the original and modified models

Features	Statistically significant at p < 0.05	
	Original model 1	Modified model 1
aindian	yes	no
black	yes	no
disable	yes	no
hosp	yes	no
vcrime	yes	no
p_liquor	yes	no
lowbirthw	no	yes

Features	Statistically significant at p < 0.1	
	Original model 5	Modified model 5
a5564	yes	no
black	yes	no
mincome	no	yes
disable	no	yes
drinking	no	yes
lowbirthw	no	yes
statenameFL	no	yes
statenameNE	no	yes
statenameUT	no	yes

Author: Rebeca Ansar
Date: September 17, 2022

Cfra based on deathp by January 23, 2021



Cfra based on deathp by August 15, 2022