MADA Project: Supplemental Material

Dawson Dobash

2021-12-10

This document displays some supplemental material to the analysis.

**Supplemental material for question 1**

This table is showing statistics for the simple linear regression between the predictor health care funds with the outcome being the logit transformation of the proportional difference between recovery and death rates. This is supplemented for table 1.2.

This helps describe that no predictors are good in predicting the outcome with a low R-squared.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.0482893 | 0.0256295 | 0.8814947 | 2.131057 | 0.1517811 | 1 | -55.85986 | 117.7197 | 123.0723 | 32.63538 | 42 | 44 |

This table is showing statistics for the simple linear regression between the predictor number of tests performed with the outcome being the logit transformation of the proportional difference between recovery and death rates. This is supplemented for table 1.2.

This helps describe that no predictors are good in predicting the outcome with a low R-squared.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.0554038 | 0.0495367 | 1.174692 | 9.443193 | 0.0024895 | 1 | -256.5248 | 519.0497 | 528.3309 | 222.1643 | 161 | 163 |

This table is showing statistics for the simple linear regression between the predictor location, which is the continent each country is associated with, with the outcome being the logit transformation of the proportional difference between recovery and death rates. This is supplemented for table 1.2.

This helps describe that no predictors are good in predicting the outcome with a low R-squared.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.1094214 | 0.0810591 | 1.155049 | 3.857979 | 0.0025218 | 5 | -251.7256 | 517.4512 | 539.1075 | 209.4596 | 157 | 163 |

This table shows the correlation matrix used to determine which variables I will use for the step-wise selection in question 1. We see that value which is the health care funds and the gdp per capita are very correlated. Since health care funds has a lot of missing, I decided to use gdp per capita.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | prop\_diff\_recov\_vs\_death | prop\_vacc | pct\_cases | test\_per\_person | Value | gdp\_per\_capita |
| prop\_diff\_recov\_vs\_death | 1.0000000 | -0.0803651 | 0.3442886 | 0.0772749 | -0.2350707 | -0.2869304 |
| prop\_vacc | -0.0803651 | 1.0000000 | 0.1684084 | 0.2318597 | 0.5676869 | 0.5756738 |
| pct\_cases | 0.3442886 | 0.1684084 | 1.0000000 | 0.1536646 | 0.1593376 | 0.0140531 |
| test\_per\_person | 0.0772749 | 0.2318597 | 0.1536646 | 1.0000000 | 0.2479436 | 0.2898869 |
| Value | -0.2350707 | 0.5676869 | 0.1593376 | 0.2479436 | 1.0000000 | 0.8735441 |
| gdp\_per\_capita | -0.2869304 | 0.5756738 | 0.0140531 | 0.2898869 | 0.8735441 | 1.0000000 |

This table is showing the process of the stepwise selection for the logit transformed difference in proportions. The nvmax means the number of predictors used in each model and each row shows the best model from each quantity of variables used. Although the number of variable shows 12, it is only 6 distinct variables. I used the package caret and the function I used considers each level of nominal variables as a separate variable. This mean the location variable and the government variable is split up by the levels. We can see that nvmax value 3 has the lowest RMSE value which mean it is the best model to use. We can also see that the R-squared values are all very low, which is somethin I addressed in the manuscript.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| nvmax | RMSE | Rsquared | MAE | RMSESD | RsquaredSD | MAESD |
| 1 | 1.039660 | 0.0537335 | 0.7644332 | 0.1705357 | 0.0486793 | 0.0799745 |
| 2 | 1.047080 | 0.0557990 | 0.7628809 | 0.1597340 | 0.0846832 | 0.0907926 |
| 3 | 1.026351 | 0.0784807 | 0.7656666 | 0.1562872 | 0.1022312 | 0.0849125 |
| 4 | 1.069087 | 0.0548963 | 0.7780939 | 0.1707713 | 0.0806150 | 0.0748228 |
| 5 | 1.076030 | 0.0377686 | 0.7840597 | 0.2052413 | 0.0563367 | 0.0886395 |
| 6 | 1.062123 | 0.0460462 | 0.7892126 | 0.1536465 | 0.0524189 | 0.0787061 |
| 7 | 1.090377 | 0.0278254 | 0.8019227 | 0.1929947 | 0.0337856 | 0.1013284 |
| 8 | 1.093651 | 0.0207601 | 0.8113988 | 0.1903827 | 0.0317245 | 0.0939165 |
| 9 | 1.084103 | 0.0273487 | 0.8028117 | 0.1954675 | 0.0312963 | 0.0977827 |
| 10 | 1.083214 | 0.0301785 | 0.7995457 | 0.1901623 | 0.0337828 | 0.0939732 |
| 11 | 1.093521 | 0.0220567 | 0.8104572 | 0.1884856 | 0.0312768 | 0.0930608 |
| 12 | 1.093521 | 0.0220567 | 0.8104572 | 0.1884856 | 0.0312768 | 0.0930608 |

**Supplemental material for question 2**

This table is showing statistics for the simple linear regression between the predictor percentage of cases with the outcome being the logit transformation of the proportion of people vaccinated.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.3424685 | 0.3383845 | 1.541764 | 83.85519 | 0 | 1 | -300.8479 | 607.6958 | 616.977 | 382.7026 | 161 | 163 |

This table is showing statistics for the simple linear regression between the predictor GDP per capita with the outcome being the logit transformation of the proportion of people vaccinated.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.3441518 | 0.3396287 | 1.519314 | 76.08774 | 0 | 1 | -269.0611 | 544.1223 | 553.0936 | 334.7056 | 145 | 147 |

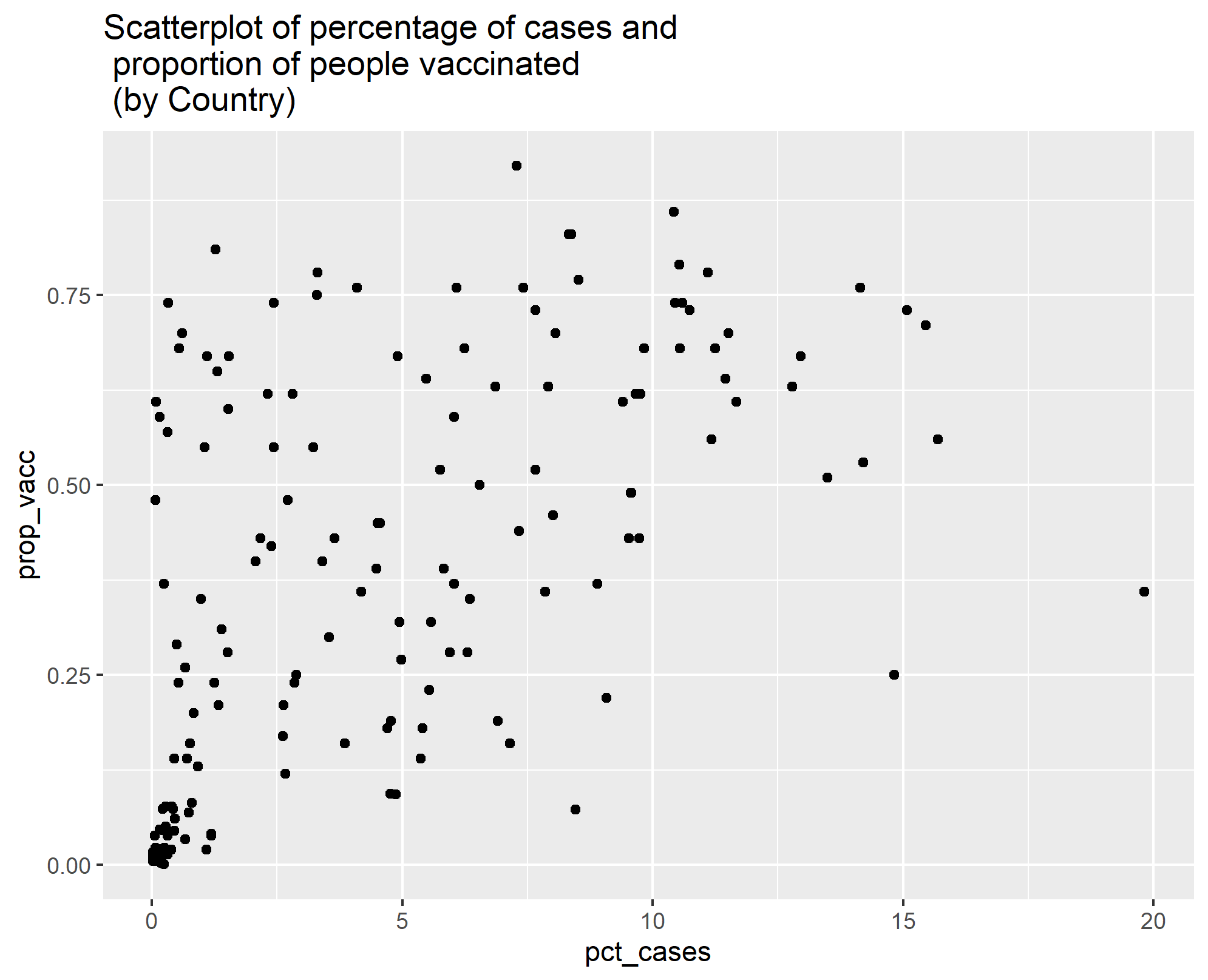
This table is showing statistics for the simple linear regression between the predictor number of COVID-19 tests per person with the outcome being the logit transformation of the proportion of people vaccinated.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.1779365 | 0.1728306 | 1.7239 | 34.84863 | 0 | 1 | -319.0489 | 644.0977 | 653.379 | 478.4651 | 161 | 163 |

This table is showing statistics for the simple linear regression between the predictor location which is the continent with the outcome being the logit transformation of the proportion of people vaccinated.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r.squared | adj.r.squared | sigma | statistic | p.value | df | logLik | AIC | BIC | deviance | df.residual | nobs |
| 0.479069 | 0.4624789 | 1.389673 | 28.8767 | 0 | 5 | -281.8686 | 577.7371 | 599.3934 | 303.1971 | 157 | 163 |

This figure shows that there is no trend between vaccinations and cases until you get out to over 10% of the population had cases of COVID-19. The vaccination numbers range a lot when there is a low perecentage of cases compared to the countries population.



This figure shows the tuning process in a visual. The best model is represented by the lowest RMSE within the dip.

