Sean Rhee

Eriksson

ECN 140

10 December 2023

COVID 19 Race Gender Poverty Risk

Abstract:

The intention of this dataset (COVID-19 Race Gender Poverty Risk - U.S. Counties) was to understand the relationship between race and ethnicity, gender, poverty and severe health conditions caused by Covid 19. The primary question motivating this research is: How has poverty status evolved within the last year, and what factors contribute to these changes? To address this, we examined the effect of income, education, employment, sex, age, and race on the percentage of people under or above the poverty line. The findings reveal significant shifts in poverty status across various demographic groups. We identify that economic hardship is a persistent issue for certain populations, moreover, the study presents education and employment status as key determinants of poverty risk. These results underline the necessity of targeted policy interventions to address poverty inequality and emphasize the importance of opportunities in combating poverty in the United States.

Introduction:

This study aims to highlight the issue of COVID-19 with poverty and other variables. Poverty has been an issue that has affected many countries, especially developing economies. Many countries have challenged the issue of poverty, but it has been an unsuccessful attempt. Although the United States is a developed country with high GDP rates and success in exports and imports, we as a country have conflict with dealing with our poverty rates. This data teaches us what variables have a positive or negative correlation on the poverty line and what policies or opportunities we can open up to alleviate poverty in the United States.

In this project, we examine the dataset to address the following key questions: What factors influence poverty status? Which variables alleviate poverty or cause poverty? What are some steps that we can take by analyzing these variables' effect on the poverty line? We employed an Ordinary Least Squares (OLS) regression analysis to estimate the relationship between poverty status and income, education attainment, work experience, race origin, and other factors.

Our analysis of the dataset reveals many economic phenomena. We found that despite overall economic growth in GDP and other indicators of the economy, poverty remains a persistent challenge for certain demographic groups, emphasizing the importance of targeted policy interventions. Education, work experience, and employment status highlight the essentialness of education and job opportunities in combating economic hardship and poverty. These 3 factors emerge as significant determinants of poverty risk. The data offers valuable insights into the economy of the United States and underscore the urgency of addressing poverty-related disparities in order to alleviate poverty for those who are struggling against it and raise our economic status as a country.

Reference/Contribution:

The *National Library of Medicine* addresses the effect of COVID-19 on home health beneficiaries. COVID-19 has had a huge effect on the global economy, especially in the United States. The impact of COVID-19 has been devastating to people of all ethnicities and races. *NIH* concluded that infection rates, poverty levels, and death rates have been prominent in non-white communities and areas. The study also makes it prominent that high poverty levels lead to high infection and death rates. Poor neighborhoods often end up with the highest infection rates and death and there is evidence from the data that states that those who live in poor neighborhoods are often non-white and mostly African-American and hispanic. Through this information, we want to analyze why are non-white communities and poverty correlated with one another and what are some steps we can take in order to prevent a future global pandemic from crashing economies, cause tremendous high infection and death rates, and how to prevent our minority communities from being impacted by infections and mortality.

Data:

We are studying *COVID 19 Race Gender Poverty Risk (U.S. Counties),* a study gathered from information by the U.S. Census, New York Times, and USA Facts. This study is a pooled cross-sectional study as the data was gathered in 2022 and focuses on individual factors that affect a much bigger trend. The reason why it's a pooled cross section data is because it combines individual and independent cross-sectional data over a period of time. The number of observations or people analyzed was found to be 3,142 observations. The observations were then divided by their respective U.S. county and state. Each variable that was created for this study has divided the number of observations into averages and the *Federal Information Processing Standard Publication (FIPS)*. The variables we are given will help us analyze how COVID-19 cases cause such high poverty rates, mortality rates, and infection rates in ethnicities that are non-white. Through the information of the *National Library of Medicine, The U.S. Census,* and the other data references, we can predict which areas will be greatly affected by COVID-19. We can see which counties in the United States are affected the most and what groups of people will be facing the most challenging aspects of COVID-19.

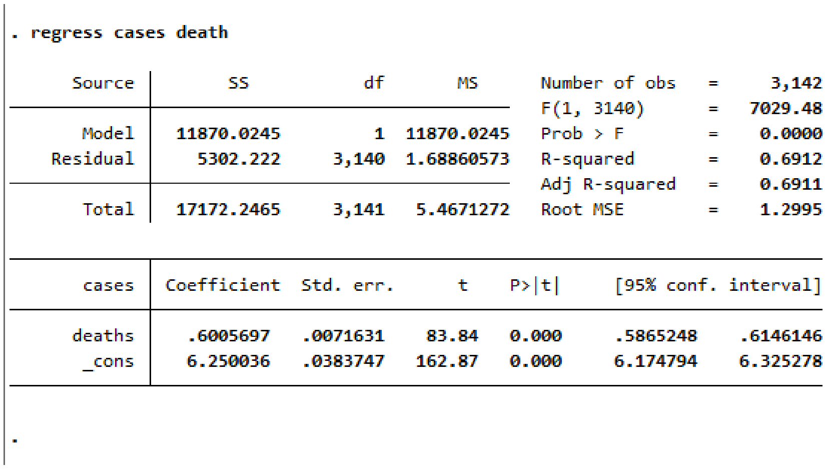
Empirical Specification:

In order to create a regression model, we need to define each variable that is provided. The primary variable we will be analyzing is the number of cases of COVID-19. Other variables that were taken into account were *County, State, Deaths, Poverty, Population, Male and Female divided into White, Black, Indian, Asian, and Non-Hispanic.* We will be analyzing the effect of the *Cases* variable with these other variables through linear regression. We will be studying and predicting how COVID cases affect deaths and poverty.

Our regression model would follow the format: $Y= \beta0 + \beta1Xi + U$. Y would represent our dependent variable, in this case, it is the case of COVID-19. B0 and β1 would represent our linear component. Xi represents the independent variable which would be *Cases of Covid 19.* U represents the random error term. With our multiple regression table, we would regress black males and females along with the poverty caused by COVID-19 and the number of deaths caused by the virus. In Stata, we would regress these variables through the command *regress cases* and our independent variables. Our second regression table would be *Regress cases deaths poverty population w_female, b_female.*

Results:

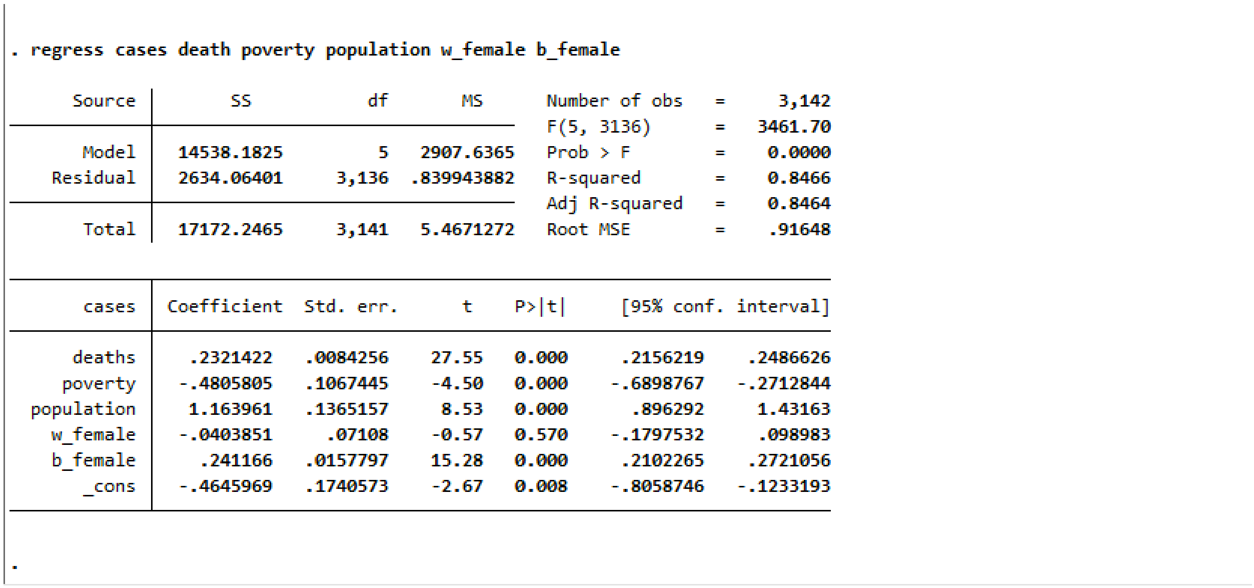
1. *Regress case deaths***:**



```
. regress cases death

      Source |       SS           df       MS      Number of obs   =     3,142
-------------+----------------------------------   F(1, 3140)      =   7029.48
       Model |  11870.0245         1  11870.0245   Prob > F        =    0.0000
    Residual |    5302.222     3,140  1.68860573   R-squared       =    0.6912
-------------+----------------------------------   Adj R-squared   =    0.6911
       Total |  17172.2465     3,141   5.4671272   Root MSE        =    1.2995

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      deaths |   .6005697   .0071631    83.84   0.000     .5865248    .6146146
       _cons |   6.250036   .0383747   162.87   0.000     6.174794    6.325278
------------------------------------------------------------------------------

.
```

For every percentage point increase in death caused by COVID-19 , there is an increase in the number of cases by 0.600pp, AEE.

2. *<u>Regress cases deaths poverty population w_female, b_female.</u> Running a nonlinear functional form causes a better fit for the data and captures non-linear effects and avoids omitted variable bias.*



```
. regress cases death poverty population w_female b_female

      Source |       SS           df       MS      Number of obs   =     3,142
-------------+----------------------------------   F(5, 3136)      =   3461.70
       Model |  14538.1825         5  2907.6365   Prob > F        =    0.0000
    Residual |  2634.06401     3,136  .839943882   R-squared       =    0.8466
-------------+----------------------------------   Adj R-squared   =    0.8464
       Total |  17172.2465     3,141  5.4671272   Root MSE        =    .91648

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      deaths |   .2321422   .0084256    27.55   0.000     .2156219    .2486626
     poverty |  -.4805805   .1067445    -4.50   0.000    -.6898767   -.2712844
  population |   1.163961   .1365157     8.53   0.000      .896292     1.43163
    w_female |  -.0403851     .07108    -0.57   0.570    -.1797532     .098983
    b_female |    .241166   .0157797    15.28   0.000     .2102265    .2721056
       _cons |  -.4645969   .1740573    -2.67   0.008    -.8058746   -.1233193
------------------------------------------------------------------------------

.
```

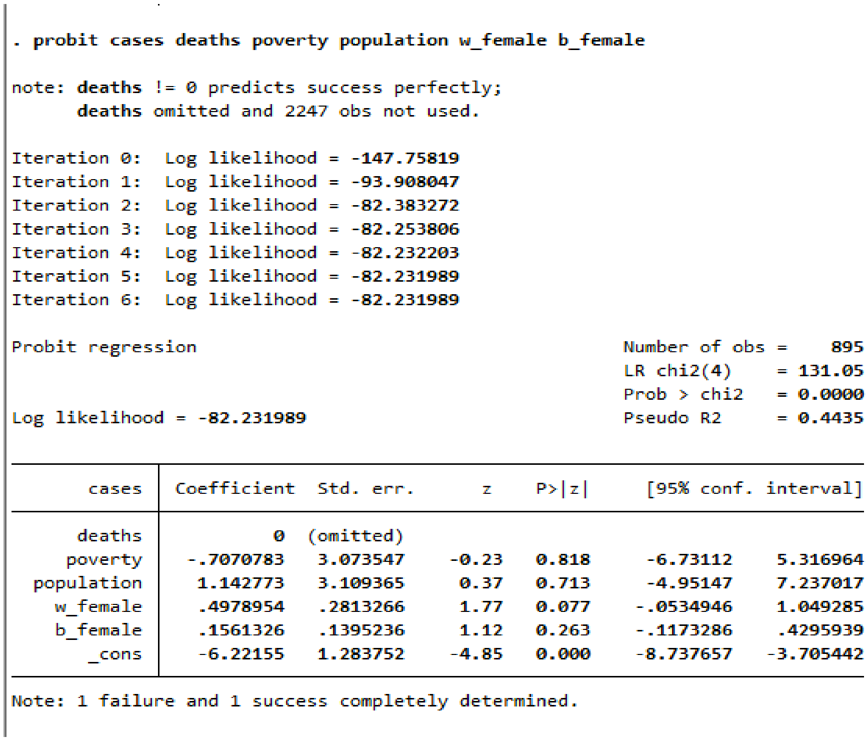
3. *The main parameter of interest is predicting how cases will affect our independent variables.*
    a. *Poverty: for every case in COVID-19, there is a 48% decrease in the individuals living in poverty. AEE.*
    b. *Population: for every case in COVID-19, there is a 1.16pp increase in the number of residents in the county with COVID-19. AEE.*
    c. *W_female: for every case in COVID-19, there is a 4.03pp decrease in the number of white females in the county facing poverty . AEE.*
    d. *B_female: There is a .24pp increase in the number of black females in the county with COVID-19 facing poverty. AEE.*

4. *Regress cases deaths*: α = 0.05. P-score = 0.00. Fail to reject the null hypothesis because p-score < 0.05

5. *Regress cases deaths poverty population w_female, b_female: All fail to reject the null hypothesis except for white_female. W_female: we reject the null hypothesis because the pscore was higher than 0.05. P-score: 0.570. 0.570 > 0.05*

6. *Some potential omitted variable bias issues with OLS in this dataset are time-variance, correlations with the included variables. With these bias issues, this would cause our bias to go negative. In order to fix this, we can use the new data and implement instrumental variables, a new model, or even including the omitted variables.*

7. *Using a fixed effects regression or logistic/probit model is more suited than OLS because the variables are binary or limited dependent variables. An OLS model when the models are dependent but continuous and only when all the assumptions are met. Fixed effects use panel data or time-invariant variables, which this dataset uses.*



```
. probit cases deaths poverty population w_female b_female

note: deaths != 0 predicts success perfectly;
      deaths omitted and 2247 obs not used.

Iteration 0:  Log likelihood = -147.75819
Iteration 1:  Log likelihood = -93.908047
Iteration 2:  Log likelihood = -82.383272
Iteration 3:  Log likelihood = -82.253806
Iteration 4:  Log likelihood = -82.232203
Iteration 5:  Log likelihood = -82.231989
Iteration 6:  Log likelihood = -82.231989

Probit regression                                       Number of obs =    895
                                                        LR chi2(4)    = 131.05
                                                        Prob > chi2   = 0.0000
Log likelihood = -82.231989                             Pseudo R2     = 0.4435

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      deaths |          0  (omitted)
     poverty |  -.7070783   3.073547    -0.23   0.818     -6.73112    5.316964
  population |   1.142773   3.109365     0.37   0.713     -4.95147    7.237017
    w_female |   .4978954   .2813266     1.77   0.077    -.0534946    1.049285
    b_female |   .1561326   .1395236     1.12   0.263    -.1173286    .4295939
       _cons |   -6.22155   1.283752    -4.85   0.000    -8.737657   -3.705442
------------------------------------------------------------------------------
Note: 1 failure and 1 success completely determined.
```

8. *We have a positive effect with our variables on cases after using a probit model. The only variable with a negative effect is poverty. Within the probit model, when cases of COVID-19 are accounted for, the number of individuals in poverty decrease with a -7.70pp.*

Conclusion:

With our regressions and results, we can predict what can occur with COVID-19 cases and how it might affect other variables. By using the variable *Cases,* we can regress it with variables such as death, poverty, and population to see the effects of COVID-19. In this case, we used *Cases* as our dependent variable to further predict and hypothesize relationships. We also tested the statistical significance between two variables through hypothesis testing as shown in the results.

In our results, we are shown that many of the variables have a positive or negative correlation on cases. Having a negative correlation signifies that there is a negative relationship between the independent variable and the dependent variable. We see this *death, poverty, and cases* as all 3 have negative correlations. When we are testing our hypotheses for regressions, by using $\alpha = 0.05$, we can either reject the null hypothesis or fail to reject the null hypothesis. All the variables failed to reject the null hypothesis except for *poverty.* This means we have enough evidence to reject the hypothesis and conclude that we can favor the alternative hypothesis.

References:

1. *https://www.kaggle.com/datasets/laurindogarcia/covid-19-race-gender-poverty-risk-us-county*

2. US Coronavirus Cases (USA Facts/U.S CDC, 2020): timeseries from 22/01/2020 to 31/07/2020;
3. US Coronavirus Deaths (USA Facts/U.S. CDC, 2020): timeseries from 22/01/2020 to 31/07/2020;
4. State/County Poverty Universe Data, All ages (SAIPE, U.S Census, 2019);
5. Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019 (CC-EST2019-ALLDATA) (U.S Census, 2019);
6. Severe COVID-19 Health Risk Index by U.S County (Policy Map/NY Times/2017 SMART-BRFSS, U.S CDC, 2017)
7. Cabin, William. "Pre-Existing Inequality: The Impact of Covid-19 on Medicare Home Health Beneficiaries." *Home Health Care Management & Practice*, SAGE Publications, May 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC7871046/#:~:text=Overall%2C%20they%20found%20that%20the,higher%20than%20in%20wealthier%20neighborhoods.

Appendix:

Figures:

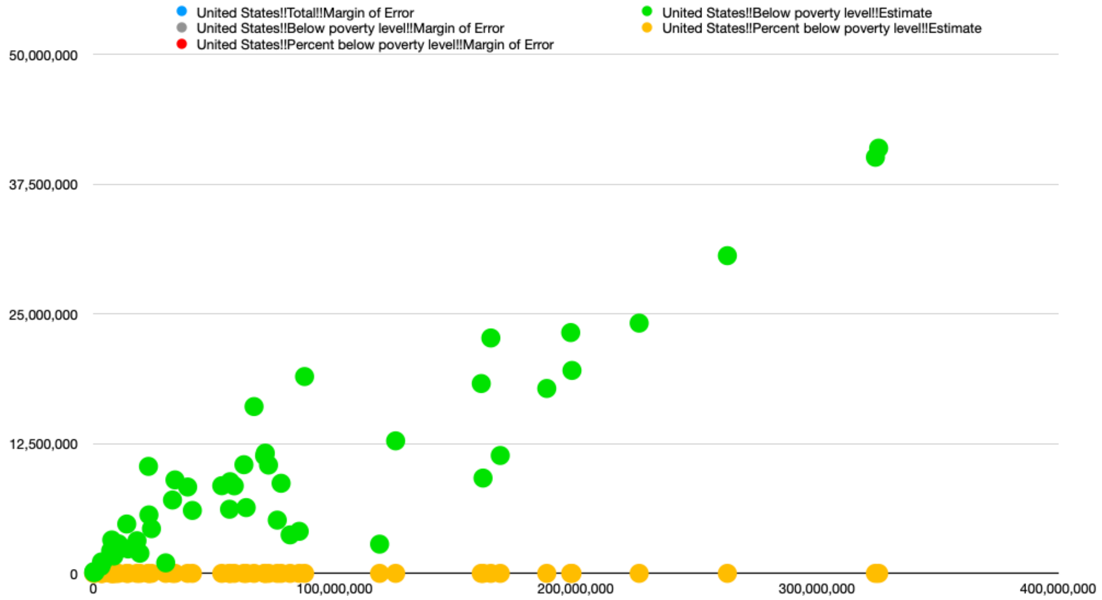
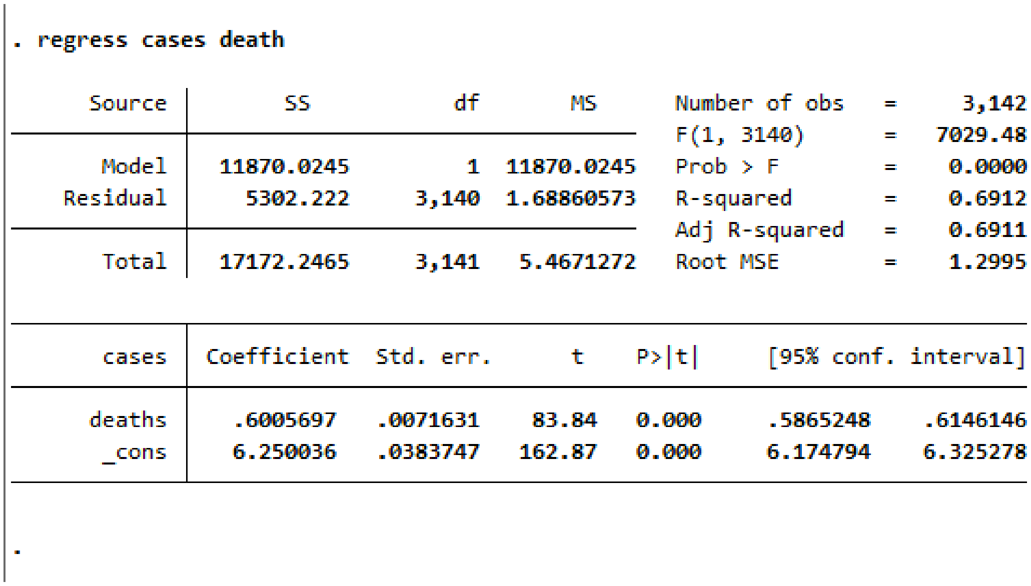


Figure 1: a scatter plot of the regression.

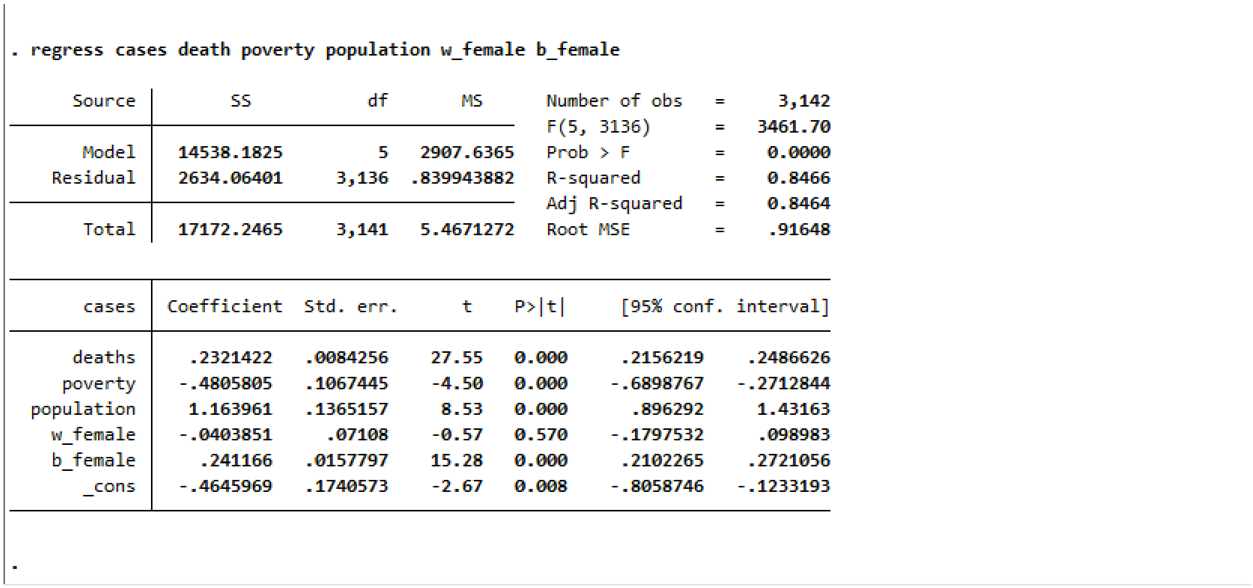
Tables:

Table 1: Summary of all the variables, observations, mean, standard deviation, minimum, and maximum of the covid_data_log.



```
. regress cases death

      Source |       SS           df       MS      Number of obs   =     3,142
-------------+----------------------------------   F(1, 3140)      =   7029.48
       Model |  11870.0245         1  11870.0245   Prob > F        =    0.0000
    Residual |    5302.222     3,140  1.68860573   R-squared       =    0.6912
-------------+----------------------------------   Adj R-squared   =    0.6911
       Total |  17172.2465     3,141   5.4671272   Root MSE        =    1.2995

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      deaths |   .6005697   .0071631    83.84   0.000     .5865248    .6146146
       _cons |   6.250036   .0383747   162.87   0.000     6.174794    6.325278
------------------------------------------------------------------------------

.
```

Table 2: The regression between cases of COVID-19 and the deaths caused by COVID-19.



```
. regress cases death poverty population w_female b_female

      Source |       SS           df       MS      Number of obs   =     3,142
-------------+----------------------------------   F(5, 3136)      =   3461.70
       Model |  14538.1825         5  2907.6365   Prob > F        =    0.0000
    Residual |  2634.06401     3,136  .839943882   R-squared       =    0.8466
-------------+----------------------------------   Adj R-squared   =    0.8464
       Total |  17172.2465     3,141  5.4671272   Root MSE        =    .91648

------------------------------------------------------------------------------
       cases | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      deaths |   .2321422   .0084256    27.55   0.000     .2156219    .2486626
     poverty |  -.4805805   .1067445    -4.50   0.000    -.6898767   -.2712844
  population |   1.163961   .1365157     8.53   0.000      .896292     1.43163
    w_female |  -.0403851     .07108    -0.57   0.570    -.1797532     .098983
    b_female |    .241166   .0157797    15.28   0.000     .2102265    .2721056
       _cons |  -.4645969   .1740573    -2.67   0.008    -.8058746   -.1233193
------------------------------------------------------------------------------

.
```

Table 3: a regression table between cases on death, poverty, population, white females, and black females.