Finding the best predictor to evaluate the issue of Suicide

Sha Jameson

May 2024

Abstract

Suicide is death caused by injuring oneself with the intent to die. Over the years, more than ever, it has become a significant public health concern worldwide. According to the CDC [Centers for Disease Control and Prevention, 2024b], Suicide rates has increase 36% between 2000 - 2022. This has prompted many different governments and organizations to help raise awareness and make prevention efforts. However, to make efforts in suicide prevention, the cause(s) must be identified This paper examins the potential predcitors of suicide rates by looking at variables including alcohol consumption, education levels, GDP per capita, hours worked per week, and internet access.

Keywords

- Suicide
- Suicide prevention
- Mental health
- COVID-19 impact
- Alcohol consumption
- Internet usage
- Economic factors

Introduction

Mental health has become an increasingly important issue for the general public. For many years, people's well-being has been overlooked, and we have been asked to put aside our own needs for the public good. This became even more apparent during the COVID-19 pandemic. While staying at home and isolating was necessary for everyone's overall wellness, the unforeseen negative effects of the extended lockdown became evident. The population experienced an increase in mental health crises, a surge in substance abuse including alcohol, a rise in internet usage, and food consumption. Moreover, the lockdown also resulted in a decrease in what humans need, including human interactions, recreational activities, and access to family and friends, which are all imperative for maintaining human fulfillment.

Data

The data for this study has been sources from creditable sources including the US Government databases, the World Bank and the World Health Organization. As the times of each variable can vary, the data are all based on 2018, 2019, 2020 depending on sufficiency of data and its completeness. The data that has been collected can very accurately show if there is a correlation between its predictor variable and suicide.

Methods

This study employs multiple regression analysis and ANOVA to investigate the relationship between suicide rates and predictor variables such as GDP per capita, alcohol consumption, education levels, hours worked, and internet accessibility. Each predictor will be evaluated for its statistical significance and contribution to the model, using p-values and confidence intervals to determine the robustness of the associations.

Simulation

Simulations will be conducted to test the robustness of our findings under various hypothetical scenarios. Using Monte Carlo simulations, we will project future trends in suicide rates based on current data, examining how changes in economic conditions or public health strategies might alter these trends. This approach will help in understanding the potential impact of interventions and economic recovery post-pandemic.

Application

The findings from this study will be crucial for policymakers, healthcare providers, and community leaders seeking to develop targeted interventions for suicide prevention. By understanding the key predictors of suicide, strategies can be tailored to address the specific needs and risk factors identified, enhancing the effectiveness of public health campaigns and support services.

[Centers for Disease Control and Prevention, 2024a]

[World Bank, 2024b] [Our World in Data, 2024] [World Bank, 2024a] [Central Intelligence Agency, 2024]

Summary statistics

```
##
## Call:
## lm(formula = SR ~ alcohol, data = official)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7450 -2.3707 -0.9724  1.6884 14.4908
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.1742     1.2961   4.764 1.52e-05 ***
## alcohol       0.3932     0.1586   2.480   0.0164 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.153 on 53 degrees of freedom
## Multiple R-squared:  0.104,  Adjusted R-squared:  0.08705
## F-statistic: 6.149 on 1 and 53 DF,  p-value: 0.01636
```

```
##
## Call:
## lm(formula = SR ~ education, data = official)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3314 -2.4665 -0.4957  1.9514 16.0330
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.2900     3.1480   1.680   0.0988 .
## education     0.3885     0.3177   1.223   0.2268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.326 on 53 degrees of freedom
## Multiple R-squared:  0.02744,    Adjusted R-squared:  0.009088
## F-statistic: 1.495 on 1 and 53 DF,  p-value: 0.2268
```

```
##
## Call:
## lm(formula = SR ~ gdp, data = official)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9663 -2.6876 -0.6266  1.9145 14.9538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.383e+00  8.852e-01   9.470 5.45e-13 ***
## gdp         2.429e-05  2.339e-05   1.038    0.304
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.343 on 53 degrees of freedom
## Multiple R-squared:  0.01994,    Adjusted R-squared:  0.001449
## F-statistic: 1.078 on 1 and 53 DF,  p-value: 0.3038


##
## Call:
## lm(formula = SR ~ work, data = official)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1109 -2.8377 -0.5919  1.9934 15.7956
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.043922   4.246805   3.778 0.000402 ***
## work        -0.003775   0.002278  -1.657 0.103455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.277 on 53 degrees of freedom
## Multiple R-squared:  0.04925,    Adjusted R-squared:  0.03131
## F-statistic: 2.745 on 1 and 53 DF,  p-value: 0.1035


##
## Call:
## lm(formula = SR ~ internet, data = official)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3527 -2.6237 -0.4647  2.0920 15.0783
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.83684    1.97431   2.956  0.00464 **
## internet     0.04602    0.02686   1.713  0.09247 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.27 on 53 degrees of freedom
## Multiple R-squared:  0.05249,    Adjusted R-squared:  0.03461
## F-statistic: 2.936 on 1 and 53 DF,  p-value: 0.09247
```

The data summary above gives a good overview on the models. The only significant predictor we see here is Alcohol consumption with a p-val $= 0.0164$ which is < 0.05. All of the other predictors are not significant.

Correlation coefficent

Alcohol consumption vs Suicide rate

```
## [1] 0.3224218
```

Education vs Suicide rate

```
## [1] 0.1656453
```

GDP vs Suicide rate

```
## [1] 0.1412101
```

Hours worked vs Suicide rate

```
## [1] -0.2219136
```

Internet usage vs Suicide rate

```
## [1] 0.2291049
```

Looking at the Correlation coefficents, we can see that non of the variable have a correlation stronger than 0.5 or -0.5 meaning there is very low or close to no correlation with the highest being 0.3224218 between Alcohol consumption vs Suicide rate

Anova

```
## Analysis of Variance Table
##
## Response: SR
##           Df Sum Sq Mean Sq F value  Pr(>F)
## alcohol    1 106.03 106.032  6.1489 0.01636 *
## Residuals 53 913.94  17.244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

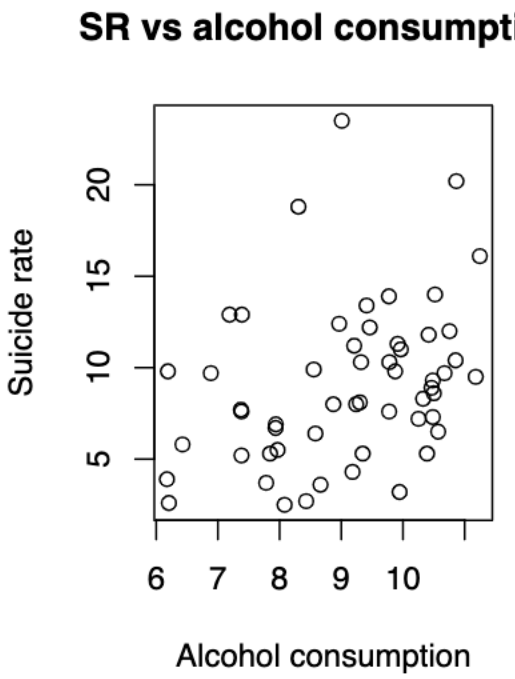
```
## Analysis of Variance Table
##
## Response: SR
##           Df Sum Sq Mean Sq F value Pr(>F)
## education  1  27.99  27.986  1.4953 0.2268
## Residuals 53 991.98  18.717
```

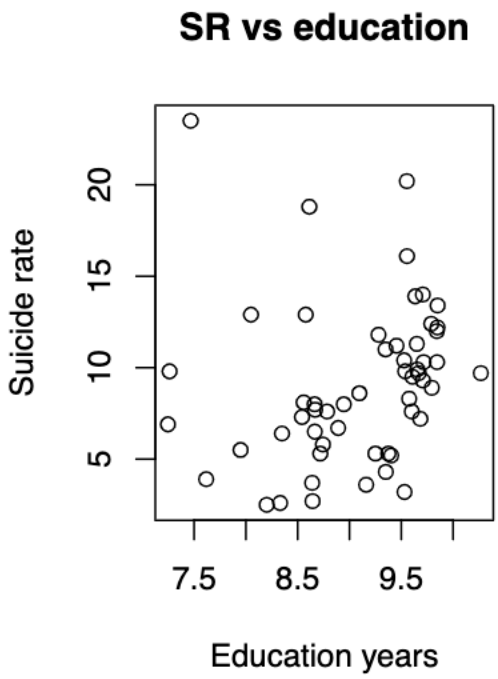
```
## Analysis of Variance Table
##
## Response: SR
##           Df Sum Sq Mean Sq F value Pr(>F)
## gdp        1  20.34  20.338  1.0783 0.3038
## Residuals 53 999.63  18.861
```

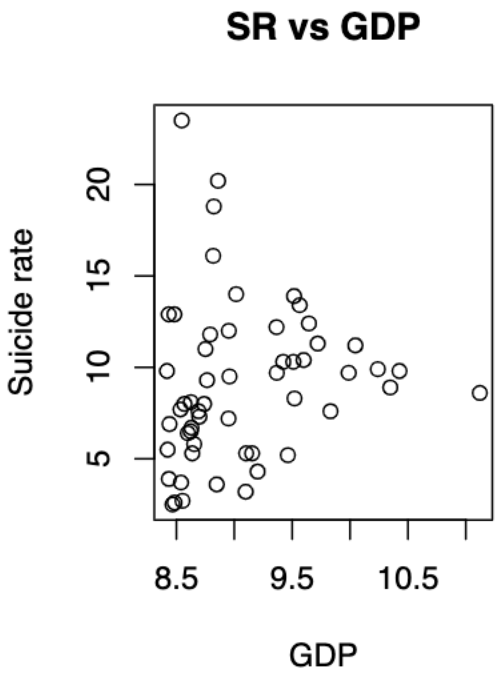
```
## Analysis of Variance Table
##
## Response: SR
##           Df Sum Sq Mean Sq F value Pr(>F)
## work       1  50.23  50.229  2.7452 0.1035
## Residuals 53 969.74  18.297
```

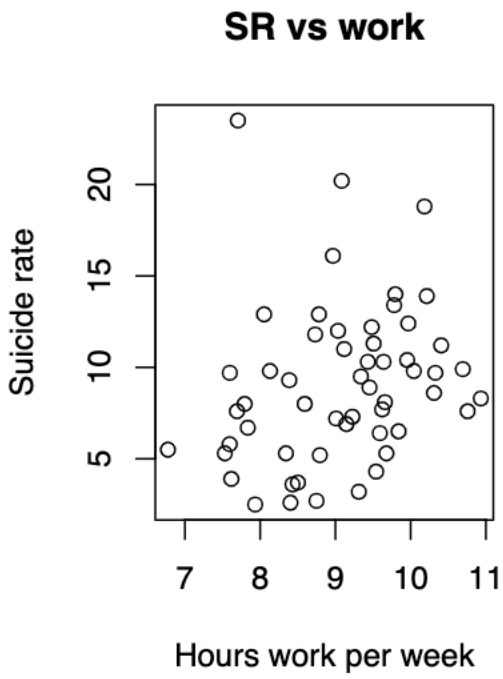
```
## Analysis of Variance Table
##
## Response: SR
##           Df Sum Sq Mean Sq F value  Pr(>F)
## internet   1  53.54  53.537   2.936 0.09247 .
## Residuals 53 966.43  18.235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

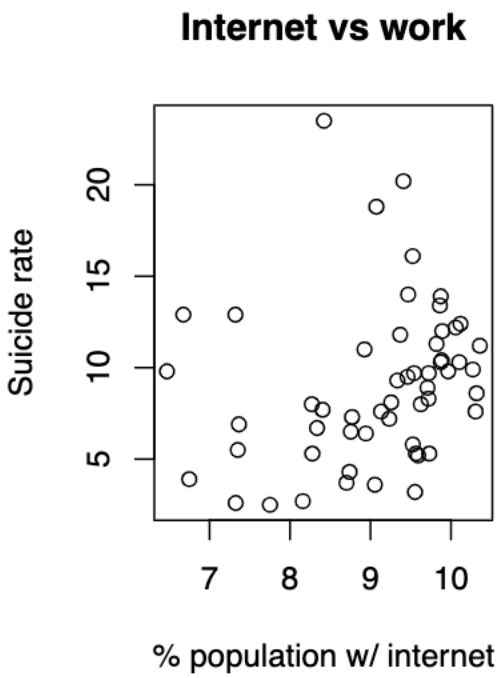
From the above ANOVA tests, we can see that none of the f-value are large and none of the p-value are $<$ 0.05 indicating that non of the variables are significant.


SR vs alcohol consumption
Suicide rate
5
10
15
20
6
7
8
9
10
Alcohol consumption


SR vs education
Suicide rate
5
10
15
20
7.5
8.5
9.5
Education years


SR vs GDP
Suicide rate
5
10
15
20
8.5
9.5
10.5
GDP


SR vs work
Suicide rate
5
10
15
20
7
8
9
10
11
Hours work per week



The plots above show little to no correlation between the predictor and its dependent variable. SR vs alcohol consumption , SR vs education and SR vs Internet usage show a slight positive linear correlation but it is faint.

```
## Summary for comb 1 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## alcohol      1  106.0  106.03   6.036 0.0174 *
## gdp          1    0.4    0.42   0.024 0.8777
## Residuals   52  913.5   17.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Summary for comb 2 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## alcohol      1  106.0  106.03   6.049 0.0173 *
## work         1    2.5    2.49   0.142 0.7080
## Residuals   52  911.5   17.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Summary for comb 3 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## alcohol      1  106.0  106.03   6.040 0.0174 *
## internet     1    1.1    1.11   0.063 0.8025
## Residuals   52  912.8   17.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Summary for comb 4 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## alcohol      1  106.0  106.03   6.036 0.0174 *
## gdp          1    0.4    0.42   0.024 0.8777
## Residuals   52  913.5   17.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Summary for comb 5 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## education    1     28  27.986   1.470  0.231
## gdp          1      2   1.953   0.103  0.750
## Residuals   52    990  19.039
##
## Summary for comb 6 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## education    1   28.0   27.99   1.507  0.225
## work         1   26.4   26.42   1.423  0.238
## Residuals   52  965.6   18.57
##
## Summary for comb 7 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## education    1   28.0   27.99   1.510  0.225
## internet     1   28.4   28.37   1.531  0.222
## Residuals   52  963.6   18.53
##
## Summary for comb 8 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## gdp          1   20.3   20.34   1.091  0.301
## work         1   30.0   29.96   1.607  0.211
```

```
## Residuals   52  969.7   18.65
##
## Summary for comb 9 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## gdp          1   20.3   20.34   1.096  0.300
## internet     1   34.8   34.85   1.878  0.176
## Residuals   52  964.8   18.55
##
## Summary for comb 10 :
##             Df Sum Sq Mean Sq F value Pr(>F)
## work         1   50.2   50.23   2.733  0.104
## internet     1   14.0   13.95   0.759  0.388
## Residuals   52  955.8   18.38
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## alcohol      1  106.0  106.03   5.746 0.0204 *
## education    1    1.0    1.03   0.056 0.8140
## gdp          1    2.3    2.26   0.122 0.7281
## work         1    1.8    1.78   0.096 0.7575
## internet     1    4.6    4.60   0.249 0.6199
## Residuals   49  904.3   18.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at our two way Anova analysis, we can see that most two way ANOVA did not improve the significantce of the data with the exception of education with internet usage, gdp with hours worked, and gdp with internet usage. However their respective p-value of 0.222, 0.211, 0.176, it is still not smaller than 0.05 meaning we cannot reject null. This indicates the statistics is not significant nor does it show any clear correlation.

With all of the variables together, the p-val kept getting larger which shows by adding each variable, it decreased it significance.

Discussion and conclusion

The results of this study indicate that there is no significant correlation between suicide rates and any of the independent variables examined.

The study faced several limitations, primarily due to insufficient data and the omission of potentially influential variables, such as the availability or accessibility of mental health care. [World Health Organization, 2024] Moreover, as the perception of mental health and suicide is still evolving, with significant stigma attached in various countries, comprehensive data on this subject remain scarce.

Ultimately, this research underscores the complexity of suicide, which does not stem from a singular cause. Various factors contribute to suicide, and these factors can vary significantly across different cultures and regions. Therefore, to effectively lower suicide rates, it is essential for each country and even specific regions within countries to consider a broad spectrum of influencing factors.

Acknowledgements

I would like to express my gratitude to Professor Yuwen Gu for his guidance and encouragement.

[Allaire et al., 2024] [**?**] [Xie, 2015] [Wickham, 2023] [Xie, 2024]

References

JJ Allaire, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. *rmarkdown: Dynamic Documents for R*, 2024. URL https://github.com/rstudio/rmarkdown. R package version 2.26, https://pkgs.rstudio.com/rmarkdown/.

Centers for Disease Control and Prevention. Mental health: Learn more, 2024a. URL https://www.cdc.gov/mentalhealth/learn/index.htm. Accessed: 2024-05-01.

Centers for Disease Control and Prevention. Suicide facts, 2024b. URL https://www.cdc.gov/suicide/facts/index.html. Accessed: 2024-05-01.

Central Intelligence Agency. Alcohol consumption per capita (country comparison), 2024. URL https://www.cia.gov/the-world-factbook/field/alcohol-consumption-per-capita/country-comparison. Accessed: 2024-05-01.

Our World in Data. Working hours, 2024. URL https://ourworldindata.org/working-hours. Accessed: 2024-05-01.

Hadley Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*, 2023. URL https://stringr.tidyverse.org. R package version 1.5.1, https://github.com/tidyverse/stringr.

World Bank. Gdp per capita (current us$) - popular indicators, 2024a. URL https://databank.worldbank.org/indicator/NY.GDP.PCAP.CD/1ff4a498/Popular-Indicators. Accessed: 2024-05-01.

World Bank. Internet users (% of population) - indicator, 2024b. URL https://data.worldbank.org/indicator/IT.NET.USER.ZS. Accessed: 2024-05-01.

World Health Organization. Mental health, 2024. URL https://www.who.int/health-topics/mental-health#tab=tab_1. Accessed: 2024-05-01.

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL https://yihui.org/knitr/. ISBN 978-1498716963.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2024. URL https://yihui.org/knitr/. R package version 1.46.