

Students and COVID-19

Original Dataset Name: “COVID-19 and its impact on education, social life and mental health of students: A Survey”

- Curators: Aya Hussein, Callie Sullivan, Maryam Jameel
- Source:
<https://www.kaggle.com/datasets/kunal28chaturvedi/covid19-and-its-impact-on-students>
- Principal Investigators: Kunal Chaturvedi, Dinesh Vishwakarma, Nidhi Singh
- Social Science Field(s): Psychology, Education, Developmental Psychology
- Number of Files: 1
- Sample Size: 1182
- Time Period: 7/13/2020 - 7/17/2020
- Method of Collection: Online Survey
- Population of Interest: Students (any age)
- Location: India
- Unit of Observation: Individuals

Description

This dataset focuses on the impact of the COVID-19 lockdown on students in India. Responses came from students of different ages from educational institutions in the Delhi National Capital Region (NCR). It explores topics such as health, diet, and screen time. It contains multiple binary and interval numeric variables, as well as categorical variables. It also includes two variables from free response questions, providing more personalized responses from participants. This presents researchers with an opportunity to code these responses in as granular or aggregated a manner as they deem appropriate.

This dataset may be of particular interest to educational professionals since it focuses on the effectiveness of online classes on students. It is also a particular interest to child psychology researchers on the mental effects of lockdown and online classes on students.

Potential Uses

- Teaching students statistical analysis techniques
- Teaching students to interpret data
- Example of survey research methodology
- Correlation analysis
- Linear regression, bivariate and multivariate
- Logistic regression, bivariate and multivariate

Sample Analyses

Did the type of device that students used for online class affect their attitude towards online class?

```
df$devices[df$devices=="NA"] <- NA
freq(df$devices)
df$device.any<- as.numeric((df$devices=="Any Gadget")+(df$devices=="Smartphone or
Laptop/Desktop"))
df$device.cpu<- as.numeric(df$devices=="Laptop/Desktop")
df$device.phone<- as.numeric(df$devices=="Smartphone")
df$device.tablet<- as.numeric(df$devices=="Tablet")
freqC(df$device.any)

logit_ortg.device = glm(df$ortg.likert~
      +df$device.phone
      +df$device.tablet
      +df$device.any) #device.cpu is reference category

summary(logit_ortg.device)
```

	Estimate	Std. Error	P-value
(Intercept)	2.6716	0.0581	0.0000
Phone	-0.1409	0.0825	0.0877
Tablet	0.3825	0.2306	0.0975
Any Device	0.8284	0.4332	0.0561

Were youth who used video games to de-stress less likely to experience health problems during the lockdown?

#REQUIRES TRANSFORMATION

#How do we take multiple text responses and turn them into usable material?

#Video Game vs. Not

```
freq(df$destress)
df$gamers <- as.numeric(df$destress=="Online gaming")
freq(df$gamers)
```

```
summary(glm(df$lockdown_health.probs=="YES" ~ df$gamers))
```

#Gamers are 6% less likely to have reported having a health problem during the lockdown (HA!
All those hours alone in front of the computer means you aren't out catching diseases, of course!)

	Estimate	Std. Error	P-value
(Intercept)	0.1450	0.0108	0.0000
Gamers	-0.0593	0.0281	0.0349

What if we control for age? After all, younger people play games more; maybe age is the real reason, since older people are more likely to report health problems

```
summary(glm(df$lockdown_health.probs=="YES" ~ df$gamers+df$age))
```

#age doesn't account for it

	Estimate	Std. Error	P-value
(Intercept)	0.0982	0.0386	0.0112
Gamers	-0.0554	0.0282	0.0498
Age	0.0023	0.0018	0.2075

Variables Included (see Codebook)

Numeric - Binary

more_connected	Do you find yourself more connected with your family, close friends , relatives ? Yes/NO
lockdown_health.probs	Did you have any health issues during lockdown? Yes/No
time utilized	Time utilized? Yes/NO
region	Delhi-NCR/Outside

	Delhi (Region of Residence)
time_well.spent	Are you satisfied with your utilization of time during lockdown? Yes/No

Numeric - Interval

age	Age of Subject
online.class_hours	Time Spent on Online Class
online.class_rating	Rating of Online Class Experience excellent/poor
self.study_hours	Daily hours spent on self study
fitness_hours	Daily hours spent on fitness
sleep_hours	Daily hours spent on sleep
soc.media_hours	Daily hours spent on social media
tv_hours	Daily hours spent watching TV
daily_meals	Meals eaten per day

Categorical

online.class_device	Medium For Online Class
weight_change	Change in your weight
soc.media_platform	Preferred Social Media Platform

Free Response

destress	How did you cope with rising stress levels
lockdown_missing	What did you miss most during lockdown

Installation

Before you begin, make sure you have the following installed:
RStudio

Downloading the Data

[We don't have a formalized process for this yet, but put instructions here. The file is called "df.csv"]

Importing the Data into RStudio

Thanks to the University of Wisconsin School of Business for providing the basis for these instructions!
For more, check them out by [clicking here](#)

1. Click on the *Environment* tab in the upper right window of RStudio
 - a. Next, click the *Import Dataset* tab
 - b. Choose *From Text (readr)*...
2. Click *Browse...* to open the files saved on your device
 - a. Find *df.csv* wherever you saved the data file; double click to open
3. A window will pop up with your dataset. On the lower left is a box labeled *Import Options*
 - a. Tick the box that says *First Row As Names*
 - b. In the box to the right of the word *Delimiter*, open the menu and select *comma*