Self-Reported Mental Health of Americans 2015-2021

Amanda Doty
DS240, Spring 2023

The Question

How does attention and the ability to concentrate relate to Americans' perception of their mental health?

I am interested in this question because I am personally studying my own mental and physical health. I am interested in the correlation between my hyperfocus days and the way I feel, as well as those days when my focus flits from one topic to the next. This survey asked thousands of Americans their perceptions of their own mental health every year from 2015 to 2021.

The Data Set

What is in the data?

The data consist of:

- location_id (state or territory identification) - 385 non-null int64 values
- year - 385 non-null int64 values
- DECIDE_yes_value (Had serious difficulty concentrating as a percent of survey sample) - 385 non-null float64 values
- DECIDE_yes_sample (Had serious difficulty concentrating individual answers) - 385 non-null int64 values
- DECIDE_no_value (Did not have serious difficulty concentrating as a percent of the survey sample) - 385 non-null float64 values
- DECIDE_no_sample (Did not have serious difficulty concentrating individual answers) - 385 non-null int64 values

- resp222_value (Zero days of poor mental health as a percent of survey sample) - 164 non-null float64 values
- resp222_sample (Zero days of poor mental health individual answers) - 164 non-null float64 values
- resp223_value (1-13 days of poor mental health as a percent of the survey sample) - 164 non-null float64 values
- resp223_sample (1-13 days of poor mental health individual answers) - 164 non-null float64 values
- resp224_value (14 or more days of poor mental health as a percent of the survey sample) - 164 non-null float64 values
- resp224_sample (14 or more days of poor mental health individual answers) - 164 non-null float64 values

The Questions from the Survey - see references at end of presentation

Resp222, 223, and 224:

Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good? (pg. 17)

Answer options:

- _ _ Number of days (01- 30)
- None
- Don't know/not sure
- Refused

DECIDE_yes and DECIDE_no:

Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions? (pg. 36)

Answer options:

- Yes
- No
- Don't know / Not sure
- Refused

More about the data set

- I found the data set on kaggle, using the search term "mental health"
  - https://www.kaggle.com/datasets/asasherwyn/brfss-mental-health-2015-2021?resource=download
- The original data set is in CSV format, and includes multiple files:
  - The big data set that consists of the data from the previous slide
  - Files that define location_id and column names
- Size:
  - Big data set - 19 kb
  - Location_id key - 833 bytes
  - Columns key - 348 bytes
- Data source:
  - https://www.cdc.gov/brfss/

About BRFSS - see references at end of presentation

- The Behavioral Risk Factor Surveillance System (BRFSS) collect[s] state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.

- BRFSS collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world.

- By collecting behavioral health risk data at the state and local level, BRFSS has become a powerful tool for targeting and building health promotion activities. As a result, BRFSS users have increasingly demanded more data and asked for more questions on the survey

Exploration

Quality - see references at end of presentation

- In a telephone survey such as the BRFSS, a sample record is one telephone number in the list of all telephone numbers the system randomly selects for dialing.
- The target population (aged 18 years and older) for cellular telephone samples in 2021 consists of people residing in a private residence or college housing who have a working cellular telephone.
- In 2021, a mean of 108,839 people were contacted via landline per state and a mean of 183,914 people were contacted via cellular telephone per state.
  - Of those samples, a mean of 2,223 landline responders were eligible and willing to complete the survey (2%), and 6,081 cellular telephone responders were eligible and willing to complete the survey via cellular telephone per state (3.3%)

Shape

- The data set, before removing null rows, has 385 rows and 12 columns
- After removing null rows, the data set has 164 rows and 12 columns
  - Rows relating to the years 2015 through 2018 were removed from the set as the question about days of poor mental health were not yet included in the survey, therefore those columns had null values
  - I made the decision to remove these rows to aid in the correlation and regression functions later

Column and row details

I changed the column names to be more descriptive of what the series represents.

BEFORE:

	location_id	year	DECIDE_yes_value	DECIDE_yes_sample	DECIDE_no_value	DECIDE_no_sample	resp222_value	resp222_sample	resp223_value	resp223_sample	resp224_value	resp224_sample
0	1	2015	15.0	1174	85.0	6526	NaN	NaN	NaN	NaN	NaN	NaN
1	1	2016	14.3	979	85.7	5812	NaN	NaN	NaN	NaN	NaN	NaN
2	1	2017	15.8	998	84.2	5498	NaN	NaN	NaN	NaN	NaN	NaN
3	1	2018	15.8	958	84.2	5423	NaN	NaN	NaN	NaN	NaN	NaN
4	1	2019	16.7	1016	83.3	5763	58.0	4302.0	24.5	1580.0	17.5	1040.0

	location_id	year	Serious difficulty concentrating	Number of YES responses	No serious difficulty concentrating	Number of NO responses	Zero days of poor mental health	Zero days responses	1-13 days of poor mental health	1-13 days responses	14 or more days of poor mental health	14+ days responses
4	1	2019	16.7	1016	83.3	5763	58.0	4302.0	24.5	1580.0	17.5	1040.0
5	1	2020	14.0	687	86.0	4465	60.8	3319.0	23.4	1177.0	15.8	757.0
6	1	2021	15.1	570	84.9	3853	56.7	2748.0	24.9	1061.0	18.4	705.0
11	10	2019	10.5	393	89.5	3352	61.3	2452.0	25.3	865.0	13.5	474.0
12	10	2020	9.8	363	90.2	3435	65.2	2648.0	22.5	829.0	12.3	482.0
13	10	2021	10.8	335	89.2	3105	63.0	2311.0	24.1	830.0	13.0	430.0
18	11	2019	10.2	222	89.8	2281	54.0	1573.0	34.4	719.0	11.5	260.0
19	11	2020	10.8	271	89.2	2996	51.6	1947.0	33.9	992.0	14.5	410.0
20	11	2021	10.4	237	89.6	2773	50.5	1764.0	35.3	1020.0	14.2	347.0
25	12	2019	12.9	2178	87.1	13841	64.2	10585.0	22.0	3235.0	13.8	2279.0

After changing column titles and removing null rows (head)
New shape: (164, 12)

Analysis and Findings

Methodologies

- Originally, I approached the data with the assumption that people who struggle with concentration also report more days of poor mental health, so I performed correlation and regression tests for the series "Serious difficulty concentrating" as the independent variable, and "1-13 days of poor mental health" and "14 or more days of poor mental health" as the dependent variables.
- However, there is little correlation between the difficulty to concentrate and…anything else (except an exact -1 correlation between "Serious difficulty concentrating" and "No serious difficulty concentrating"

Methodologies - continued

- I instead found stronger correlations between "No serious difficulty concentrating" and at least 1 day of poor mental health, as seen in the heat map on the right, with the strongest being the negative correlation between 14 or more days and no difficulty concentrating
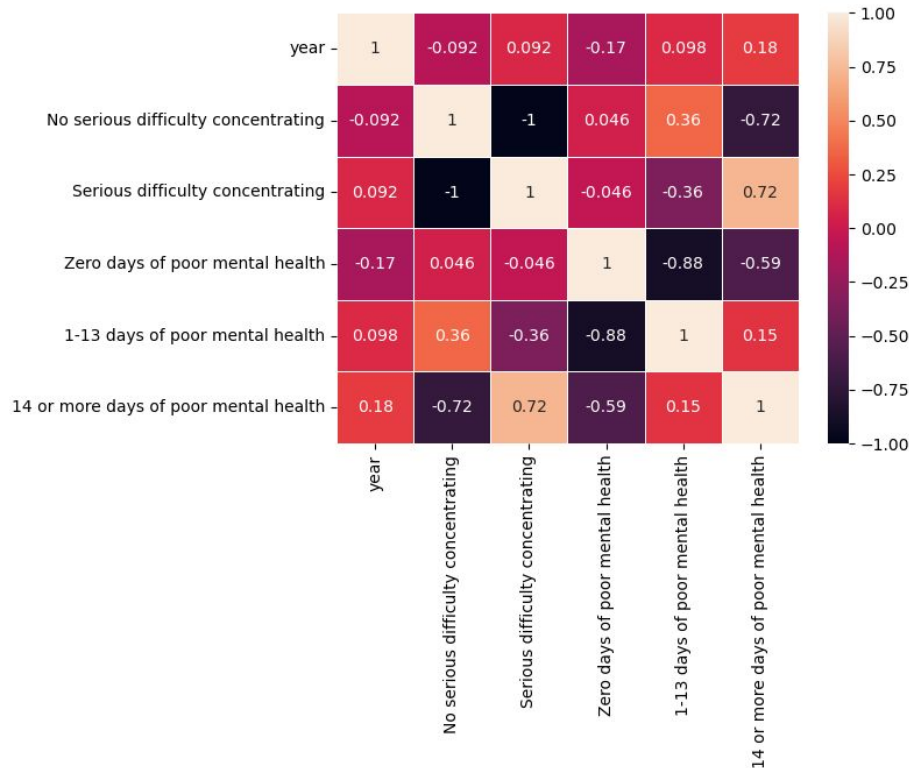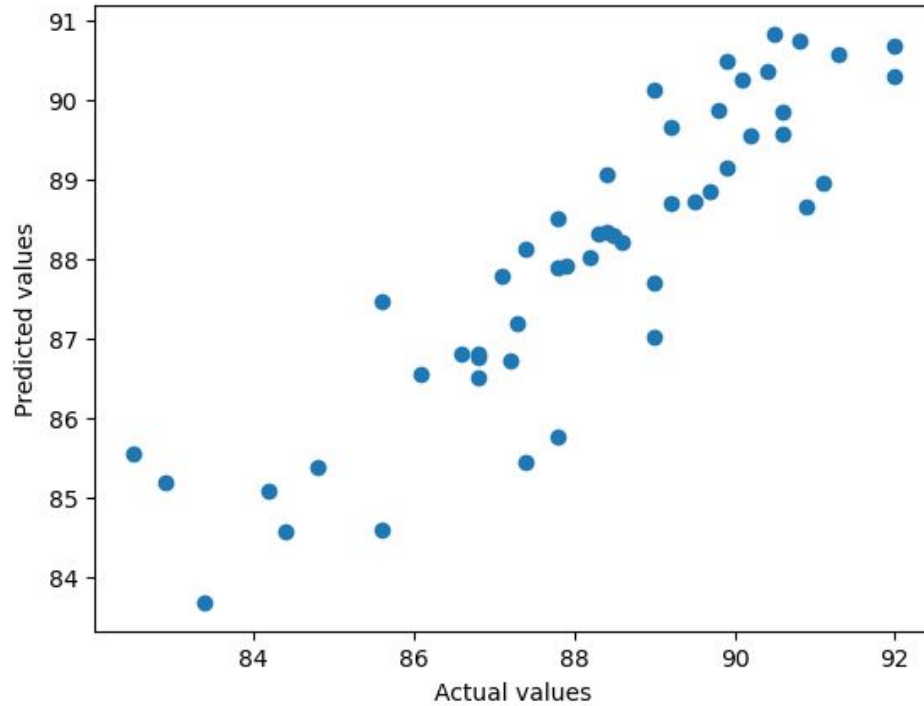
Linear Regression

x = data[['1-13 days of poor mental health','14 or more days of poor mental health']]

y = data[['No serious difficulty concentrating']]

Intercept: [93.17764086]

Coefficients: [[0.32697035 -0.94295352]]

Linear Regression - continued

Linear Regression - continued

Variance: 0.7864632406004994

R2: 0.7812266033259123

Mean squared: 1.1533742223593768

Methodologies - adding dependent variable

When I added another dependent variable, I found that the regression was less accurate. This is because the strongest correlations were opposite, one was negative and the other was positive.

Linear Regression 2

x=data[['1-13 days of poor mental health','14 or more days of poor mental health','Zero days of poor mental health']]
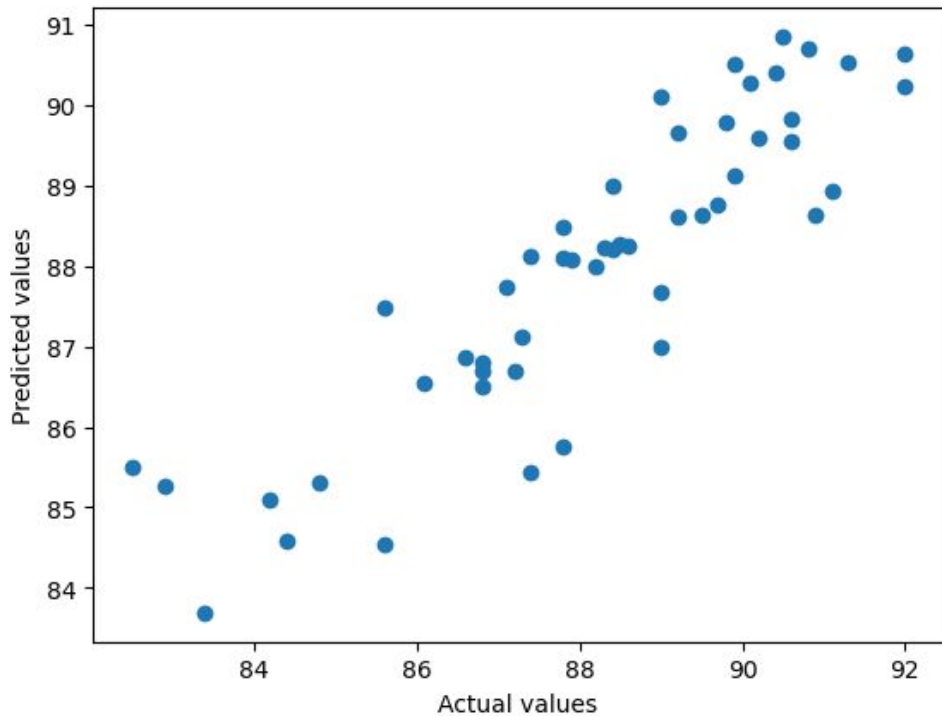
No change in Y

Intercept: [151.26491118]

Coefficient: [[-0.25433249 -1.51898527 -0.58208012]]

Linear Regression 2 - continued

Intercept: [151.26491118]

Coefficient: [[-0.25433249 -1.51898527
-0.58208012]]

Analysis

- There is a strong negative correlation between respondents reporting no difficulty concentrating and 14 or more days of poor mental health
- In other words, people who do not struggle with concentration due to illness or malady tend to have more good mental health days per month
- However, the relationship is not terribly strong, so using one to predict the other is not recommended

Conclusion

Using this data set is helpful for my own study because it shows me that there is more that goes into mental health than two variables. One may play with the other, but they do not necessarily go hand-in-hand.