

Midlife in the United States: Daily Stress Project, 2004-2009

Initial Data

The datasets curated were pulled from a published research project titled “Midlife in the United States: Daily Stress Project, 2004-2009”. In the original study, the data was obtained via an 8-day phone interview. The original researchers developed several scales in which the data could be utilized to derive meaningful information. This data was gathered from 2004 to 2009 to observe how various factors may influence stress. Participants were interviewed eight consecutive times within a year. This was not a longitudinal study. In the initial dataset, there were 16176 rows and 546 columns.

Dataset Information

Two datasets were curated using the data provided by this research project. The first (“paired_t_test_stress_score.csv”) dataset selects the year 2004 and pulls the stress variables utilized in the original study. Using the stress scale provided by the research project, the total stress score of the first interview was compared to the total stress score of the eighth interview. This data set has 652 rows and 11 columns after removing 1856 null values. The second dataset (“physical_symptoms.csv”) was an attempt to understand what factors influence stress the most. An ANOVA was utilized to observe how physical symptoms, such as muscle soreness, impact stress based on age groups. This dataset has 654 rows and 33 columns after removing 30 null values.

Data Processing & Cleaning

When creating these datasets, we utilized the codebook provided by the authors of the study. There were several values used to represent missing data, whether it was a skipped question or a non-response, the lack of responses was coded differently. The first step to creating a workable dataset was handling missing data. To calculate the total stress score, the missing values were dropped from the dataset. The large size of the dataset prevents the removal of missing values from impacting the usability of the data. No participants were excluded from this dataset unless they had not responded to all questions in this scale.

Data Use & Limitations

The primary use of this dataset was to better understand how various factors impact stress. The dataset was cleaned and processed using the guidelines from the original research study. There are potential limitations in how missing data was handled. Individuals who responded to 75% of the stress scale may have been removed when their data could have been insightful. In terms of responses to the interviews, there may have also been some bias. Many questions were asked about drug and alcohol usage. Individuals may try to appeal to what is socially desirable and alter their responses, therefore impacting the data.

Codebook – Paired T-test:

("paired_t_test_stress_score.csv")

Variable Name	Question	Scoring	Original Names
ID	Participant ID		M2ID
interview_number	Interview day (1-8)		B2DDAY
year	Interview year (2004-2009)		B2DIYEAR
argument_disagreement	Did you have an argument or disagreement with anyone since the last time we spoke?	0=No; 1=Yes	B2DF1
conflict_avoidance	Since the last time we spoke, did anything happen that you could have argued about that you let pass to avoid a disagreement?	0=No; 1=Yes	B2DF2
work_school_stress	Since the last time we spoke, did anything happen at work or school that most people would consider stressful?	0=No; 1=Yes	B2DF3
home_stress	Since the last time we spoke, did anything happen at home that most people would consider stressful?	0=No; 1=Yes	B2DF4
discrimination_experience	Many people experience discrimination on the basis of race, sex, and age, did anything like this happen to you since the last time we spoke?	0=No; 1=Yes	B2DF5
family_friend_stress	Since the last time we spoke, did anything happen to a relative or close friend that turned out to be stressful for you?	0=No; 1=Yes	B2DF6
other_stress	Did anything else happen to you that people would consider stressful?	0=No; 1=Yes	B2DF7
total_stress_score	Sum of yes values	0=No; 1=Yes	B2DN_STR

***All null values were excluded from the dataset

Codebook – ANOVA:

("physical_symptoms_anova.csv")

Variable Name	Question	Scoring	Original Names
Id	Participant id	Unique identifier	M2ID
Year	Interview year	2004-2009	B2DIYEAR
Interview_number	Interview day	1-8	B2DDAY
Age		21-40; 41-60; 60+	B1PAGE_M2
Headache	Experienced a headache?	0=No; 1=Yes	B2DB1A
Backache	Experienced a backache?	0=No; 1=Yes	B2DB1B
Muscle_soreness	Experienced muscle soreness?	0=No; 1=Yes	B2DB1C
Fatigue	Experienced fatigue?	0=No; 1=Yes	B2DB1D
Joint_pain	Experienced joint pain?	0=No; 1=Yes	B2DB1E
Muscle_weakness	Experienced muscle weakness	0=No; 1=Yes	B2DB1F
Cough	Experienced a cough	0=No; 1=Yes	B2DB1G
Sore_throat	Experienced sore throat	0=No; 1=Yes	B2DB1H
Fever	Experienced a fever?	0=No; 1=Yes	B2DB1I
Chills	Experienced chills?	0=No; 1=Yes	B2DB1J
Other_cold_or_flu_symptoms	Experienced other cold and flu symptoms?	0=No; 1=Yes	B2DB1K
Nausea	Experienced nausea?	0=No; 1=Yes	B2DB1L
Allergies	Experienced allergies?	0=No; 1=Yes	B2DB1LA
Diarrhea	Experience diarrhea?	0=No; 1=Yes	B2DB1M
Constipation	Experienced constipation?	0=No; 1=Yes	B2DB1N

Poor_appetite	Experienced poor appetite?	0=No; 1=Yes	B2DB1O
Other_stomach_problems	Experienced other stomach problems?	0=No; 1=Yes	B2DB1P
Chest_pain	Experienced chest pain?	0=No; 1=Yes	B2DB1Q
Dizziness	Experienced dizziness?	0=No; 1=Yes	B2DB1R
Shortness_of_breath	Experienced shortness of breath?	0=No; 1=Yes	B2DB1S
Menstrual_related_symptoms	Experienced menstrual related symptoms	0=No; 1=Yes	B2DB1T
Hot_flashes_or_flushes	Experienced hot flashes or flushes?	0=No; 1=Yes	B2DB1U
Any_other_physical_symptoms	Experienced any other physical symptoms?	0=No; 1=Yes	B2DB1V
Skin_related_symptoms	Experienced skin related symptoms?	0=No; 1=Yes	B2DB1W
Eye_related_symptoms	Experienced eye related symptoms?	0=No; 1=Yes	B2DB1X
Ear_related_symptoms	Experienced ear related symptoms	0=No; 1=Yes	B2DB1Y
Teeth_related_symptoms	Experienced teeth related symptoms	0=No; 1=Yes	B2DB1Z
Leg_or_foot_related_symptom	Experienced leg or foot related symptoms	0=No; 1=Yes	B2DB1LEG
Total_score	Sum of Items		B2DN_SYM

***All null values were excluded from the dataset