Final Applied Project

New Attempt

- Due Dec 15, 2024 by 11:59pm
- Points 200
- Submitting a file upload
- Attempts 2
- Allowed Attempts 3

Final Applied Project

The final applied project, based on the assigned dataset, will constitute a significant portion of your final grade in this course.

About the Data

Each student will work on an assigned dataset. You can access the dataset and related materials from the provided zip file.

The dataset originates from the **National Survey of Children's Health (NSCH)**. The NSCH supports national initiatives to enhance the health and development of children by providing national and state-level data on key measures of child health and well-being. These data are crucial for understanding the health status and healthcare needs of children across the nation and within individual states and communities.

Useful Data Resources:

- **National Survey of Children's Health (HRSA) (https://mchb.hrsa.gov/data-research/national-survey-childrens-health)**
- **Child Health Data Center (https://www.childhealthdata.org/dataset)**
- **Census Bureau Data Dictionary (https://www.census.gov/data-tools/demo/uccb/nschdict)**

Guidelines for the Project

Dataset Overview

- Each assigned zip file contains:
  - Two CSV files: one with variable names and another with data values.
  - Supporting documentation about the survey and other relevant details.
- The project focuses on analyzing the provided dataset for one of the following three health topics (outcomes):

1. **Chronic Diseases**
2. **Mental Health**
3. **Neurodevelopmental Disabilities (NDD)**

Health Topics (Outcomes)

1. **Chronic Diseases**
   Chronic diseases are broadly defined as conditions lasting one year or more, requiring ongoing medical attention, limiting daily activities, or both. Examples include heart disease, cancer, and diabetes. These conditions often involve multiple variables within the dataset.

2. **Mental Health**
   Mental health conditions in the NSCH dataset include issues such as anxiety and depression. These were assessed through parent responses to questions such as:

   - *"Has a doctor or other healthcare provider EVER told you that this child has a mental health disorder?"*
   - If "yes," the follow-up question was: *"Does this child CURRENTLY have the condition?"*
     Mental health disorders were considered present if the parent answered "yes" to the second question regarding conditions such as depression, anxiety, or ADHD.

3. **Neurodevelopmental Disabilities (NDD)**
   NDD includes conditions such as Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Anxiety, Learning Disability (LD), Speech Delay (SD), Intellectual Disability (ID), Down Syndrome (DS), Developmental Delay (DD), Seizures, and Behavioral/Mood Problems (BP).

Analysis Requirements

- Identify relevant variables in the dataset for the chosen health topic (outcome).
  - Outcomes like chronic diseases or NDD often involve multiple variables.
- Select **at least 20 independent variables (risk factors)** for analysis.
  - These variables may include:
    - **Individual-level factors:** age, gender, race/ethnicity, BMI, physical activity, time spent watching TV or on electronic devices.
    - **Clinical conditions:** comorbidities and related health factors.
    - **Family environment factors:** family income, smoking environment, neighborhood safety, and social support.
- Analyze the dataset to:
  1. Examine the associations between the characteristics of participants and outcomes.
  2. Use machine learning methods to investigate the relationship between risk factors and outcomes.

- Highlight insights into variable importance and selection within your analysis.

Documentation

1. Compile a **data dictionary** for the variables used in your project, ensuring thorough documentation for each variable.
   - Variable names can be found in the assigned zip file.
   - Detailed descriptions for variables (organized by year) can be accessed via the **Census Bureau Data Dictionary (https://www.census.gov/data-tools/demo/uccb/nschdict)** .
2. Provide:
   - Clear descriptions of each variable.
   - Information on how they were utilized in your analysis.

Instructions for Final Applied Project

General Guidance

1. **Carefully read the dataset instructions.**
   Familiarize yourself with the structure, variables, and context of your assigned dataset.
2. **Access to your dataset.**
   Each student is assigned a dataset to work with for their project.
3. **Understand your dataset.**
   Thoroughly explore the data to identify key variables and their relationships.
4. **Develop your project based on your dataset.**
   Your analysis should align with the data provided, and you may refine your scope based on your study design.
5. **Subset selection (optional):**
   If your study design requires, you may work on a selected subset of the data. However, your final working dataset must include **at least 10,000 observations (rows)** to ensure statistical robustness.

Format for Submission

Your applied project must be submitted in the format of an **academic article**, consisting of the following sections:

1. **Title**
   A concise, informative title that reflects your study's purpose and scope.

2. **Abstract**
   A brief summary of the objectives, methods, key findings, and conclusions of your study (approximately 250 words).

3. **Background**
    - Provide context and significance of the chosen topic.
    - Highlight relevant literature and motivation for your study.
4. **Study Design**
    - **Aims:** Clearly state the goals of your analysis.
    - **Population Selection:** Describe how participants/observations were chosen.
    - **Data and Materials:** Outline the dataset and relevant documentation used.
5. **Statistical Analysis**
    - **Methods:** Explain your approach, including statistical or machine learning techniques applied.
    - **Response Variable:** Define the dependent variable (outcome) for your study.
    - **Independent Variables/Risk Factors:** Detail the predictors or factors examined.
    - **Variable Selection/Model Selection:** Describe the methods used to identify significant variables and choose appropriate models.
6. **Results**
    - **Model Used:** Specify the models applied (e.g., linear regression, random forest).
    - **Model Evaluation Metrics:** Report metrics such as accuracy, R^2, AUC, or similar measures to validate your models.
    - **Major Findings:** Summarize the key results.
    - **Tables and Figures/Plots:** Include relevant tables, plots, or visualizations to illustrate your findings.
7. **Discussion**
    - **Summary of Methods and Results:** Recap the key points of your study.
    - **Strengths:** Highlight the strengths and novel contributions of your analysis.
    - **Limitations:** Discuss any weaknesses or constraints of your study design or analysis.
8. **Conclusions**
   Provide concise, actionable takeaways from your analysis.
9. **References**
   Use a consistent academic citation style to list all references cited in your project.

Supplementary Materials

In addition to the academic article, please submit the following supplementary documents:

1. **Final Analyzed Dataset**
   The subset of the data used in your final analysis.

2. **Data Dictionary**
   A comprehensive list of the variables used in your project, including:
   - Variable names.
   - Descriptions.
   - Data types.
   - Value ranges or categories.

3. **R/Python Code**
   Provide the scripts used for data cleaning, analysis, and visualization. Ensure the code is well-documented and reproducible.

4. **Notes for Submission**

- Ensure that your submission is clear, structured, and professionally formatted.
- Tables, figures, and supplementary materials should be appropriately labeled and referenced in the main text.

Steven Edelmann: Chronic diseass **2021 NSCH.zip**
(https://udel.instructure.com/courses/1805732/files/140685083?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140685083/download?download_frd=1)

Miao Li: Mental Health **2021 NSCH.zip (https://udel.instructure.com/courses/1805732/files/140685083?wrap=1)**  **(https://udel.instructure.com/courses/1805732/files/140685083/download?download_frd=1)**

Fariha Mahfuz: Neurodevelopmental disabilities
(https://udel.instructure.com/courses/1731202/grades/5343804#tab-assignments) 2022NSCH.zip
(https://udel.instructure.com/courses/1805732/files/140684863?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140684863/download?download_frd=1)

Pranjal Modi: Chronic diseass **2019 NSCH.zip**
(https://udel.instructure.com/courses/1805732/files/140685051?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140685051/download?download_frd=1)

Fatima Moeini Pour: Chronic diseass **2020 NSCH.zip**
(https://udel.instructure.com/courses/1805732/files/140685070?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140685070/download?download_frd=1)

Manning Smith: Mental Health **2022NSCH.zip**
(https://udel.instructure.com/courses/1805732/files/140684863?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140684863/download?download_frd=1)

Dana Tango: Mental Health **2022NSCH.zip**
(https://udel.instructure.com/courses/1805732/files/140684863?wrap=1) 
(https://udel.instructure.com/courses/1805732/files/140684863/download?download_frd=1)