**Statistical Consulting Project (SCP) Assignment 3**

**PUBH 630**

**Spring 2023**

**Due April 29 at 9:00AM EST**

**Dataset**

* 2019 Behavioral Risk Factor Surveillance System (BRFSS)
  + focusing on the Southeastern US
* BRFSS survey information and codebooks: <https://www.cdc.gov/brfss/annual_data/2019/pdf/codebook19_llcp-v2-508.HTML>

**SCP Assignment 3: Data Cleaning and Descriptive Analyses (60 points)**

* Description of data cleaning process (i.e., missing values, creating or collapsing categories of a variable)
* Description of any descriptive analyses that will be conducted (i.e., frequencies, histograms or bar charts, correlation between covariates)
  + Include unfilled table shells you will use as part of your descriptive analyses
* Resources: Refer to pages 13-28 of <https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/10/managing-data_pw_final_09252013.pdf> for information on the data cleaning process and pages 6-7 and 23-27 of <https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/9/creating-analysis-plan_pw_final_09242013.pdf> for information on descriptive analyses and associated table shells.
* Table shells: No maximum limit is placed on table shell page lengths but remember you are limited to 10 tables and/or figures in your final paper.
* Page length: 2 pages single spaced size 12 font maximum excluding references table shells.
* Style: Assignment should address all points in the rubric and be written in complete sentences forming logical paragraphs.
* References: You may include in-text citations and a reference list with your assignment, but it is not required. No maximum limit is placed on reference page lengths.

**SCP Assignment 3: Rubric**

1. Description of data cleaning process (40 points)

* How will outliers be assessed in the dataset of variables you selected and what steps will you take if there are outliers in your dataset? (8 points)

All are categorical variables.

* How will missing values be identified in the dataset? (8 points)
* How will missing values be dealt with in the dataset? (8 points)
* Describe what steps you will take if there are missing values in covariates of interest (i.e., gender, race/ethnicity, age, other socioeconomic or health-related characteristics). (8 points)
* Do you plan on creating or collapsing any categories of variables included in your analyses? If so, mention which variables you plan on doing this to and include the new categories you intend to create. (8 points)

Created Don’t know = missing

1. Description of descriptive analyses (10 points)

* What univariate descriptive analyses will you conduct? (5 points)

Count, count (%), pie chart, and bar chart as all variables are categorical

* What bivariate descriptive analyses will you conduct? (5 points)

Crosstab, chi-square test with dependent variable

1. Descriptive analyses table shells (10 points)
   * Provide table shells for your bivariate analyses. (10 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Stroke** | | **No stroke** | |  |
| **Variable** | **Count (n)** | **Weighted (%)** | **Count (n)** | **Weighted (%)** | **P-value for Chi-square test** |
| **(95% CI)** | **(95% CI)** |
| **Age Group** |  |  |  |  |  |
| 18-24 | 16 |  |  |  |  |
| 25-34 |  |  |  |  |  |
| 35-44 |  |  |  |  |  |
| 45-54 |  |  |  |  |  |
| 55-64 |  |  |  |  |  |
| 65 or older |  |  |  |  |  |
| Missing\* |  |  |  |  |  |