**Case Study: SAS BRFSS 10 Categorical Table 1**

This case study will reintroduce you to categorical variables and how you may use statistical procedures to investigate these types of data, display them in a “Table 1”, and develop a results section describing your Table 1. Use data from the Behavioral Health Needs Assessment Survey from 2010 to complete this case study.

The objective of this analysis is to investigate the association between **diabetes** and **BMI** after controlling for **exercise** and **gender**. The outcome variable is **diabetes** and the variable of interest (exposure) is **BMI**.

Conduct a **complete case analysis** for this objective following these guidelines:

1. Use the raw variable categorization of BMI (\_BMI4CAT)
2. Categorize gender (SEX) into a two-level variable (male=0, female=1) where male is category 1 of the raw variable and female is category 2
3. Categorize diabetes (DIABETE2) into a two-level variable (no=0, yes=1) where yes is category 1 of the raw variable and no is category 3
4. Categorize exercise (EXERANY2) into a two-level variable (no=0, yes=1) where yes is category 1 of the raw variable and no is category 2
5. For the complete case analysis, restrict your sample based on the following conditions:
   1. 18<=AGE<=99
   2. SEX: raw categories 1 and 2
   3. DIABETE2: raw categories 1 and 3
   4. EXERANY2: raw categories 1 and 2
   5. Education (EDUCA): raw categories 1-6
   6. \_BMI4CAT: raw categories 1,2 and 3
   7. General health (GENHLTH): raw categories 1-5
6. (10 pts) Using PROC FREQ, show the simple frequency tables for gender, exercise, and BMI
7. (20 pts) Using PROC FREQ, create a 2x2 contingency table for exercise by gender.
   1. Show the PROC FREQ output
   2. Show your hand calculation of the chi-square statistic for testing whether there is an association between gender and exercise.
   3. Based on your calculated chi-square value, is there an association between gender and exercise? Explain
      1. What is your null hypothesis?
      2. What p value are use assuming?
      3. What is the critical chi-square value?
   4. Show that your (1) hand calculated chi-square statistic, and (2) conclusion on the presence of an association matches that produced by PROC FREQ. Include the relevant PROC FREQ output in your answer.
8. (30 pts) Create your “Table 1” for this objective. You can use this table template:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TABLE1. Characteristics of 413,748 BRFSS 2010 participants by BMI category.** | | | | | | | | | |
|  | **Population** | | **Normal** | | **Overweight** | | **Obese** | | **p value \*** |
| **Variable** | **N** | **%** | **n** | **%** | **n** | **%** | **n** | **%** |  |
|  |  |  |  |  |  |  |  |  |  |
| **Gender** |  |  |  |  |  |  |  |  |  |
| **Male** |  |  |  |  |  |  |  |  |  |
| **Female** |  |  |  |  |  |  |  |  |  |
| **Exercise** |  |  |  |  |  |  |  |  |  |
| **Yes** |  |  |  |  |  |  |  |  |  |
| **No** |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| \* p values based on Pearson chi-square test of association | | | | | | | | | |

1. (40 pts) Write the results section for this “Table 1”

Extra Credit (10 pts)

The calculation of a chi-square statistics makes use of an “expected value”. Using the exercise by gender contingency table, give an intuitive explanation of how the expected value is derived.