**EPI 624 Take Home Assignment 6**

The take-home assignment is worth 10 points, and is due at the end of Module 12, at 11:59 pm.

1. Please employ appropriate SAS procedures to produce descriptive statistics and p-values that would allow you to fill in the table below.
   1. Use the Dataset Assign6\_SP2021.
   2. Checking distribution of variables, data cleaning/variable creation (2 points)
      1. Calculate Age as an integer using **July 1 of this year** as the reference date.
      2. Use if/then statements to recode age from a continuous variable into a 2-level categorical variable as shown in the table below.
      3. Please recode the values of the categorical variables so that they fit appropriately in the table using if/then statements.
   3. Descriptive statistics
      1. For normally distributed continuous variables, present mean and standard deviation. For continuous variables that are not normally distributed, present median and IQR.
      2. For categorical variables, present number (count) and proportion (%).
   4. P-values
      1. If comparing means from two groups, use T-test
      2. If comparing distribution of non-normally distributed continuous variables, use non-parametric tests (e.g. Wilcoxon rank-sum test).
      3. If comparing proportions from two or more groups, choose between Chi-Square or Fisher’s Exact test, depending on expected cell sizes.
   5. Fill in the blanks for the number of observations in each age group in the table.
2. Please submit SAS code, log, and output.
   1. SAS program should have the following components:
      1. Checking distribution
      2. Producing descriptive statistics
      3. Producing p-values
3. Please also fill out the table below and include in your homework submission.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Age Group | |  |
|  | ≤ 25 Years  N = ( ) | >25 Years  N = ( ) | P-value |
| Gender, N (%)\* |  |  |  |
| Female |  |  |  |
| Male |  |  |  |
| Left Handed, N (%)\* |  |  |  |
| Yes |  |  |  |
| No |  |  |  |
| Mean Height (SD)\*\* |  |  |  |
| \*% out of total in each age group. \*\*SD: standard deviation | | | |