**HW2**

I decided to run a Kruskal-Wallis nonparametric test to determine whether a statistically significant relationship existed between the categorical risk predictor (risk) and the 12th grade alcohol use behaviors scale (alcbeh5). I decided upon this test because it could assess whether there was a statistically significant difference in the values of alcbeh5 (the dependent variable) across the three risk categories (“low”, “medium”, and “high”). It also allowed me to compare more than 2 categories of the independent variable (risk), which was not allowable for the Mann-Whitney U or the Wilcoxon signed rank tests.

The Kruskal-Wallis test assesses the null hypothesis that there is no statistical difference in the dependent variable (alcbeh5) across the three categories of the dependent variable (risk). If the p-value reported by this test is less than 0.05, then we will reject the null hypothesis at the 95% level of confidence (α = 0.05) and conclude that there is a statistically significant relationship between the two variables, alcbeh5 and risk. However, if the reported p-value is greater than or equal to 0.05, then we fail to reject the null hypothesis at the 95% level of confidence (α = 0.05) and conclude that there is *not* a statistically significant relationship between the two variables, alcbeh5 and risk.

According to the SAS® output, there is *not* a statistically significant relationship between the categorical risk variable (risk) and 12th grade alcohol behaviors (p = 0.0888 > 0.05). This verifies the results that we obtained earlier with our parametric form of regression, since our simple linear regression analysis also showed that there was *not* a statistically significant relationship between these two variables (p = 0.0994 > 0.05 for the “high” category; p = 0.4081 > 0.05 for the “medium” category). Thus, the results from both the parametric simple linear regression analysis and the nonparametric Kruskal-Wallis sensitivity analysis support the conclusion that there is *not* a statistically significant relationship between the categorical risk variable (risk) and 12th grade alcohol behaviors (alcbeh5) at the 95% level of confidence.

**SAS Code**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Course: Data Analysis - EPID 5314 \*

\* Assignment: HW 2 \*

\* Due Date: 9/9/2021 \*

\* Programmer(s): Jessie Ausman \*

\* Program Name: HW2 \*

\* Save Program/Log/Output: C:\Users\jessa\Desktop\EPID 5314\Homework2\*

\* Save Data Files: C:\Users\jessa\Desktop\EPID 5314\PNC Data File\PNC

\* Datasets \*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Read in Datasets

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

libname bios "C:\Users\jessa\Desktop\EPID 5314\PNC Data File\PNC Datasets";

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Create merged dataset

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** pnc05;

set bios.pnc05;

**run**;

**data** pnc09;

set bios.pnc09;

**run**;

**data** lab2;

merge pnc05 (in=a) pnc09 (in=b);

by ID;

if a;

if b;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Restrict set to those who drank alcohol in last year at pnc05

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **freq** data=lab2;

table alcyear4;

**run**;

**data** lab2\_1;

set lab2;

if alcyear4 NE **1**;

**run**;

**data** lab2\_2;

set lab2\_1;

if alcyear4 = **.** then delete;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Create alcohol norms variable (alc\_norm) for 8th grade data

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** lab2\_3;

set lab2\_2;

alc\_norm = frinum4 + alcsen4 + adudrnk4;

if missing(frinum4) or missing(alcsen4) or missing(adudrnk4) then alc\_norm = **.**;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Recode alc\_norm -> make it categorical

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** lab2\_4;

set lab2\_3;

if alc\_norm LE **5** then risk\_cat = "low";

if alc\_norm GE **6** and alc\_norm LE **13** then risk\_cat = "medium";

if alc\_norm GE **14** then risk\_cat = "high";

if missing(alc\_norm) then risk\_cat = " ";

else;

**run**;

\*check conditional processing;

**proc** **freq** data=lab2\_4;

tables risk\_cat\*alc\_norm;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Create dummy vars for risk\_cat

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**data** lab2\_5;

set lab2\_4;

if risk\_cat = "med" then med = **1**;

else if risk\_cat = "hig" then med = **0**;

else if risk\_cat = "low" then med = **0**;

else if missing(risk\_cat) then med = **.**;

**run**;

**data** lab2\_6;

set lab2\_5;

if risk\_cat = "med" then high = **0**;

else if risk\_cat = "hig" then high = **1**;

else if risk\_cat = "low" then high = **0**;

else if missing(risk\_cat) then high = **.**; /\*low is the reference group!\*/

**run**;

**data** lab2\_7;

set lab2\_6;

if risk\_cat = "low" then risk = **1**;

else if risk\_cat = "med" then risk = **2**;

else if risk\_cat = "hig" then risk = **3**;

else if risk\_cat = " " then risk = **.**;

else;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Run a NONPARAMETRIC test to assess the relationship between risk\_cat & alcbeh5

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*Kruskal-Wallis test;

**Proc** **npar1way** data=lab2\_7 wilcoxon dscf;

class risk;

var alcbeh5;

**run**;