**HW 10 – EPID 5314**

**Part 1:**

1. We are coding time in years and assuming that there is 1 full year between each wave of the survey. Since we code time as an increasing continuous variable, any associations between exposure and outcome observed over time will change with increasing time. Since the gap in between measurements is so long, this implies that the change in our variables over time will be stepwise since we are not actually measuring these changes in between waves. This also implies that the spacing for any graphs we create will need to be equal to 1 (since we are using 1 year intervals).

**Part 2:**

Graphical user interface

Description automatically generated

1. This graph really doesn’t tell us much since there are too many lines and the plot is too messy to truly figure out what is going on in such a large sample. The SAS log reflects this in the warning that groupmax was reached at 5,600 groups.

Chart, line chart

Description automatically generated

1. It seems that all groups increase their alcohol use over time, though Hispanics seem to have the steepest change over time as compared to the other racial groups.

1 = White

2 = Black

3 = Hispanic

4 = Other

**Part 3:**

1. Given what we have observed in our graphical analyses, it is apparent that we will need to use regression coefficients to determine if alcohol use behaviors vary over time by race. We are not interested in the overall value of the outcome in different groups because we want to determine what changed *over time*.
2. Summary statistics from the regression are as follows:

Graphical user interface

Description automatically generated

These results imply that the relationship between race and alcohol use behaviors changes over time for all races (white, Hispanic, black, and other) since the first 4 p-values are significant at the 95% level of confidence (p < 0.0001 < 0.05). Additionally, the results imply that there is a significant difference in the rate of change in alcohol behaviors between white and blacks and between whites and other races (p = 0.0002 < 0.05; p = 0.0102 < 0.05). However, there is NOT a significant difference between rate of change in alcohol behaviors between whites and Hispanics (p = 0.6915 > 0.05).

**SAS Code**

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

\* Course: Data Analysis - EPID 5314 \*

\* Assignment: Lab 9 \*

\* Due Date: 10/28/2021 \*

\* Programmer(s): Jessie Ausman \*

\* Program Name: Lab9 \*

\* Save Program/Log/Output: C:\Users\jessa\Desktop\EPID 5314\Lab10 \*

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

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

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* PART 1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

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

**data** wave1;

set bios.pnc02;

/\*create alcohol behavior cont scale var\*/

alc = alcyear1 + alcmon1 + alcweek1 + hvyuse1+ drunk1;

/\*categorize race into white, black, hisp, other\*/

if race1 = **5** then race = **1**; \*1 = white;

if race1 = **2** then race = **2**; \*2 = black;

if race1 = **3** then race = **3**; \*3 = hispanic;

if race1 in (**1**,**4**,**6**) then race = **4**; \*4 = other;

/\*create wave variable\*/

time = **0**;

**run**;

**data** wave2;

set bios.pnc03;

/\*create alcohol behavior cont scale var\*/

alc = alcyear2 + alcmon2 + alcweek2 + hvyuse2 + drunk2;

/\*categorize race into white, black, hisp, other\*/

if race2 = **5** then race = **1**; \*1 = white;

if race2 = **2** then race = **2**; \*2 = black;

if race2 = **3** then race = **3**; \*3 = hispanic;

if race2 in (**1**,**4**,**6**) then race = **4**; \*4 = other;

/\*create wave variable\*/

time = **1**;

**run**;

**data** wave3;

set bios.pnc04;

/\*create alcohol behavior cont scale var\*/

alc = alcyear3 + alcmon3 + alcweek3 + hvyuse3 + drunk3;

/\*categorize race into white, black, hisp, other\*/

if race3 = **5** then race = **1**; \*1 = white;

if race3 = **2** then race = **2**; \*2 = black;

if race3 = **3** then race = **3**; \*3 = hispanic;

if race3 in (**1**,**4**,**6**) then race = **4**; \*4 = other;

/\*create wave variable\*/

time = **2**;

**run**;

**data** wave4;

set bios.pnc05;

/\*create alcohol behavior cont scale var\*/

alc = alcyear4 + alcmon4 + alcweek4 + hvyuse4 + drunk4;

/\*categorize race into white, black, hisp, other\*/

if race4 = **5** then race = **1**; \*1 = white;

if race4 = **2** then race = **2**; \*2 = black;

if race4 = **3** then race = **3**; \*3 = hispanic;

if race4 in (**1**,**4**,**6**) then race = **4**; \*4 = other;

/\*create wave variable\*/

time = **3**;

**run**;

**data** Lab10;

set wave1 wave2 wave3 wave4;

by ID;

keep id race alc time;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* PART 2 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*Spaghetti plot - mean individual response;

**Proc** **sgpanel** data=Lab10;

panelby race / spacing = **1**;

series y = alc x = time / group = ID;

**run**;

\*Mean group response;

**Proc** **sgplot** data=Lab10;

vline time / response=alc group=race stat=mean limitstat=stderr;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* PART 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**proc** **genmod** data= Lab10;

class race (param=ref ref=first);

model alc = race time race\*time / type3;

estimate "slope for white" race **0** **0** **0** time **1**;

estimate "slope for black" race **1** **0** **0** time **1**;

estimate "slope for hispanic" race **0** **1** **0** time **1**;

estimate "slope for other" race **0** **0** **1** time **1**;

estimate "difference in slopes black vs. white" race\*time **1** **0** **0**;

estimate "difference in slopes hispanic vs. white" race\*time **0** **1** **0**;

estimate "difference in slopes other vs. white" race\*time **0** **0** **1**;

**run**;