**HW4 – EPID 5314**

1. Maximal model adjusting for all confounders to estimate the total effect

Table

Description automatically generated

1. Using change in estimate method, reduce the maximal model

**ROUND 1**

Remove race5

Graphical user interface, application

Description automatically generated with medium confidence

Remove parprob4

A picture containing table

Description automatically generated

Remove lunch4

A picture containing scatter chart

Description automatically generated

Remove alcacc\_sc5

A picture containing application

Description automatically generated

After the first round of backward elimination using the change in estimate method, the removal of parprob4 seemed to produce the least change in the exposure estimate. The change in exposure estimate produced by this removal was only 0.2%, which is much less than 10%. Thus, we will permanently remove parprob4 from the model and continue onto round 2 of backward elimination.

**ROUND 2**

Remove race5

A picture containing graphical user interface

Description automatically generated

Remove lunch4

A picture containing table

Description automatically generated

Remove alcacc\_sc5

Graphical user interface, application

Description automatically generated

After the second round of backward elimination using the change in estimate method, the removal of lunch4 seemed to produce the least change in the exposure estimate. The change in exposure estimate produced by this removal was only 0.79%, which is much less than 10%. Thus, we will permanently remove lunch4 from the model and continue onto round 3 of backward elimination.

**ROUND 3**

Remove race5

A picture containing table

Description automatically generated

Remove alcacc\_sc5

Graphical user interface, application

Description automatically generated

After the third round of backward elimination using the change in estimate method, the removal of race5 seemed to produce the least change in the exposure estimate. The change in exposure estimate produced by this removal was only 2.74%, which is much less than 10%. Thus, we will permanently remove race5 from the model and continue onto round 4 of backward elimination.

**ROUND 4**

Remove alcacc\_sc5

A picture containing application

Description automatically generated

Removal of the alcacc\_sc5 variable resulted in a 22.75% change in the exposure estimate. This percent change is much greater than 10% so we will keep alcacc\_sc5 in the model and conclude our backward elimination procedure.

1. Using the backwards elimination (one-at-a-time) method, reduce the maximal model

**MAXIMAL MODEL**

Table

Description automatically generated

**ROUND 1**

Remove parprob4 since it has the least significant p-value (p= 0.2776 > 0.2)

Table

Description automatically generated

After the first round of backwards elimination, the variable lunch4 still has an insignificant p-value (p= 0.2959 > 0.2). So, we will remove lunch4 from the model in the next round of backward elimination.

**ROUND 2**

Remove lunch4

Table

Description automatically generated

After removal of lunch4, all of the remaining p-values are significant so we are done with backwards elimination.

**CONCLUSIONS**

1. The variables I included in my maximal model were the outcome variable (atmbeh5), the dichotomized exposure variable (drunk4), race5, lunch4 (with “don’t know’ recoded as missing), parprob4, alcacc\_sc5. I excluded fridr5 from the maximal model since controlling for peer alcohol norms in 12th grade (fridr5) would have closed the indirect path in the DAG provided and only allowed us to estimate the direct effect, not the TOTAL effect.
2. Using the change in estimate method, the variables in my final model were the outcome variable (atmbeh5), the exposure variable (drunk4), and alcacc\_sc5. All other explanatory variables produced a percent change in the exposure estimate that was less than 10% when removed from the model.
3. Interpretations:

**β’s:**

Those who got drunk in the 8th grade (drunk4 = 1) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 3.9621 units greater than those who did not get drunk in the 8th grade (drunk4 = 0).

The value of the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) is expected to decrease by 0.7367 units for every unit increase in the ‘alcohol access in the 12th grade’ continuous scale (alcacc\_sc5).

**95% Confidence Intervals:**

Those who got drunk in the 8th grade (drunk4 = 1) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 3.1809 and 4.7432 units greater than those who did not get drunk in the 8th grade (drunk4 = 0).

The value of the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) is expected to decrease by between 0.8280 and 0.6453 units for every unit increase in the ‘alcohol access in the 12th grade’ continuous scale (alcacc\_sc5).

1. The variables that were included in my final model for the backwards elimination method were my outcome variable (atmbeh5), my exposure variable (drunk4), race5, and alcacc\_sc5. This was slightly different from the results I obtained from the change in estimate method of elimination.
2. Interpretation:

**β’s:**

Those who got drunk in the 8th grade (drunk4 = 1) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 3.826 units greater than those who did not get drunk in the 8th grade (drunk4 = 0).

Those who are Black (race5 = 2) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 1.3095 units greater than those who are Asian (race5 = 1).

Those who are Hispanic (race5 = 3) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 2.3848 units greater than those who are Asian (race5 = 1).

Those who are Native American (race5 = 4) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 0.9196 units greater than those who are Asian (race5 = 1).

Those who are White (race5 = 5) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 3.4061 units greater than those who are Asian (race5 = 1).

Those who are ‘other’ (race5 = 6) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is 3.5172 units greater than those who are Asian (race5 = 1).

The value of the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) is expected to decrease by 0.7232 units for every unit increase in the ‘alcohol access in the 12th grade’ continuous scale (alcacc\_sc5).

**95% Confidence Intervals:**

Those who got drunk in the 8th grade (drunk4 = 1) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 3.0502 and 4.6018 units greater than those who did not get drunk in the 8th grade (drunk4 = 0).

Those who are Black (race5 = 2) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 0.1630 and 2.4559 units greater than those who are Asian (race5 = 1).

Those who are Hispanic (race5 = 3) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 1.1979 and 3.5718 units greater than those who are Asian (race5 = 1).

Those who are Native American (race5 = 4) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between -1.7134 and 3.5526 units greater than those who are Asian (race5 = 1). (Since negative values are included in the 95% confidence interval, it is possible that Native Americans may have an atmbeh5 value that is lower than that of Asians).

Those who are White (race5 = 5) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 2.1461 and 4.6662 units greater than those who are Asian (race5 = 1).

Those who are ‘other’ (race5 = 6) are expected to have a value for the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) that is between 1.9415 and 5.0930 units greater than those who are Asian (race5 = 1).

The value of the ‘alcohol and drug use in 12th grade’ continuous scale (atmbeh5) is expected to decrease by between 0.8136 and 0.6327 units for every unit increase in the ‘alcohol access in the 12th grade’ continuous scale (alcacc\_sc5).

**BONUS:** The confounders that are selected into the final model by the significance backwards elimination approach largely depend on the significance level being used to determine deletion/selection of the confounders. For example, while some variables may be deemed invalid for the model at alpha = 0.2, increasing or decreasing the value of alpha has the potential to change the confounders included in the final model. Additionally, the significance testing approach has low power for detecting true confounders. On the other hand, the change in estimate approach does not rely on the level of significance when determining which confounders should be included in the final model. Thus, the variables included in the final models determined by each of these methods may very likely vary.

**SAS Code**

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

\* Course: Data Analysis - EPID 5314 \*

\* Assignment: HW4 \*

\* Due Date: 9/22/2021 \*

\* Programmer(s): Jessie Ausman \*

\* Program Name: HW4 \*

\* Save Program/Log/Output: C:\Users\jessa\Desktop\EPID 5314\Homework4\*

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

\* Datasets \*

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

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* PART 0 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*read in data;

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

**data** pnc09;

set bios.pnc09;

**run**;

**data** pnc05;

set bios.pnc05;

**run**;

**data** HW4;

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

by ID;

if a;

if b;

**run**;

**data** hw4\_1;

set hw4;

if sproc5 = **1**;

**run**;

\*dichotomize drunk4 into yes/no;

**data** hw4\_2;

set hw4\_1;

if drunk4 eq **1** then drunk = **0**;

else if drunk4 in (**2**,**3**,**4**,**5**,**6**) then drunk = **1**;

else if missing(drunk4) then drunk = **.**;

else;

**run**;

**proc** **freq** data=hw4\_2;

table drunk\*drunk4;

**run**;

\*set "dont know" responses for lunch4 to missing values;

**data** hw4\_3;

set hw4\_2;

if lunch4 eq **3** then lunch4 = **.**;

else;

**run**;

**proc** **freq** data=hw4\_2;

table lunch4;

**run**;

**proc** **freq** data=hw4\_3;

table lunch4;

**run**;

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

\*specify the maximal model to assess the TOTAL (direct+indirect) effect of getting drunk in the 8th

grade on alcohol and drug use in the 12th grade;

\*the maximal model should account for the following confounders:

-race5

-lunch4

-parprob4

-alcacc\_sc5

\*exposure variable:

-drunk4 --> binary

\*outcome variable

-atmbeh5 --> continuous

\*we leave out the fridr5 variable because if we were to adjust for this variable we would be estimating the direct

effect of drunk4 on atmbeh5 but not the TOTAL effect;

**proc** **genmod** data=hw4\_3;

class race5 (param=ref ref='1') parprob4 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 parprob4 lunch4 alcacc\_sc5/ type3;

**run**;

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

\*use change in estimate method to reduce the maximal model;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 1\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*remove race5;

**proc** **genmod** data=hw4\_3;

title "remove race5";

class parprob4 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk parprob4 lunch4 alcacc\_sc5/ type3;

**run**;

\*remove parprob4;

**proc** **genmod** data=hw4\_3;

title "remove parprob4";

class race5 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 lunch4 alcacc\_sc5/ type3;

**run**;

\*remove lunch4;

**proc** **genmod** data=hw4\_3;

title "remove lunch4";

class race5 (param=ref ref='1') parprob4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 parprob4 alcacc\_sc5/ type3;

**run**;

\*remove alcacc\_sc5;

**proc** **genmod** data=hw4\_3;

title "remove alcacc\_sc5";

class race5 (param=ref ref='1') parprob4 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 parprob4 lunch4/ type3;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 2\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*remove race5;

**proc** **genmod** data=hw4\_3;

title "remove race5";

class lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk lunch4 alcacc\_sc5/ type3;

**run**;

\*remove lunch4;

**proc** **genmod** data=hw4\_3;

title "remove lunch4";

class race5 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 alcacc\_sc5/ type3;

**run**;

\*remove alcacc\_sc5;

**proc** **genmod** data=hw4\_3;

title "remove alcacc\_sc5";

class race5 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 lunch4/ type3;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 3\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*remove race5;

**proc** **genmod** data=hw4\_3;

title "remove race5";

class drunk (param=ref ref='0');

model atmbeh5 = drunk alcacc\_sc5/ type3;

**run**;

\*remove alcacc\_sc5;

**proc** **genmod** data=hw4\_3;

title "remove alcacc\_sc5";

class race5 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 / type3;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 4\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*remove alcacc\_sc5;

**proc** **genmod** data=hw4\_3;

title "remove alcacc\_sc5";

class drunk (param=ref ref='0');

model atmbeh5 = drunk / type3;

**run**;

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

\*use backwards elimination (p-value) method to reduce the maximal model;

\*maximal model;

**proc** **genmod** data=hw4\_3;

class race5 (param=ref ref='1') parprob4 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 parprob4 lunch4 alcacc\_sc5/ type3;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 1\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*eliminate parprob4 since it is least significant in the maximal model;

**proc** **genmod** data=hw4\_3;

title "remove parprob4";

class race5 (param=ref ref='1') lunch4 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 lunch4 alcacc\_sc5/ type3;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*ROUND 2\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*remove lunch4 since it has least significant p-value in round 1;

**proc** **genmod** data=hw4\_3;

title "remove lunch4";

class race5 (param=ref ref='1') drunk (param=ref ref='0');

model atmbeh5 = drunk race5 alcacc\_sc5/ type3;

**run**;