Using a Factor Analysis When Exploring Survey Data

Deanna Schreiber-Gregory, North Dakota State University

Abstract

This paper looks at using a factor analysis to identify and define latent mental health and risk behavior variables in a survey data set. The study looks at recent health trends and behavior analyses of youth in America. Data used was provided by the Center for Disease Control and Prevention and gathered using the Youth Risk Behavior Surveillance System (YRBSS). A series of logistic regression analyses were then performed using the risk behavior and demographic variables as potential contributing factors to each of the mental health variables. Mental health variables included disordered eating and depression/suicidal ideation data while the risk behavior variables included smoking, consumption of alcohol and drugs, violence, vehicle safety, and sexual behavior data. Risks and benefits of using a factor analysis with logistic regression in social science research will be discussed in depth. Results included reporting differences between the years of 1991 and 2011. Data was analyzed using SAS 9.3.

Introduction

The Youth Risk Behavior Surveillance System (YRBSS) was developed as a tool to help monitor priority risk behaviors that contribute substantially to death, disability, and social issues among American youth and young adults in today's society. The YRBSS has been conducted biennially since 1991 and contains survey data from national, state, and local levels. The national Youth Risk Behavior Survey (YRBS) provides the public with data representative of the United States high school students. The state and local surveys provide data representative of high school students in states and school districts who also receive funding from the CDC through specified cooperative agreements. The YRBSS serves a number of different purposes. The system was originally designed to measure the prevalence of health-risk behaviors among high school students. It was also designed to assess whether these behaviors would increase, decrease, or stay the same over time. An additional purpose for the YRBSS is to examine the co-occurrence of different health-risk behaviors. Considering this last purpose, this particular study examines the co-occurrence of suicidal ideation, depression, and eating deregulations as indicators of psychological unrest with other health-risk behaviors. The purpose of this study is to serve as an exercise in correlating two different groups of variables across multiple years with large data sets. This paper also touches on the interactions between the health-risk behaviors and the similarities within each behavior set.

Methods

Subjects and Procedure

Participants in this study were high school students in grades 9-12 who were attending a randomly selected American High school during the year of survey administration (every other year 1991-2011). Parental permission to participate in survey is obtained prior to administration. A trained data collector was then sent to the site to administer the questionnaire. The data collector was charged with the tasks of following scripted protocol and recording information about the school and classrooms during administration. For data collection, one survey, created by CDC for that year, was administered per protocol to the students in paper format on a computer-scannable booklet or answer sheet. The questionnaire was completed in one sitting. Students absent the day of the administration were able to complete the questionnaire when they return. Collaborative efforts between CDC and Macro International ensured that proper data-processing protocols are also followed.

Inputting Data into SAS®

YRBSS provided data sets free to the public online and instructions on how to download the data sets, as well as how to apply the formatting. In order to apply the formatting, the researcher needed only to specify libraries for the data sets and formats:

```
libname mydata 'D:\SUGGF 2014';
      /* Tells SAS where the data is */
libname library 'D:\SUGGF 2014';
      /* Tells SAS where the formats are */
```

This enabled SAS® to read all the formatting as well as output the variable names, questions, and answers in a very clean manner

Data Cleaning Procedure

After inputting all the data and their formats for each year into SAS®, the next step was to clean up the data and narrow it down to only the variables that were relevant to this study. Since this study involved multiple years, the first step in data cleaning was to comb through all the years and organize the different questions and variety of answers into a reference document. Given that the survey was given in paper format, it seemed necessary to use a Microsoft Excel® workbook for this initial step. This way, the different steps in the process of reorganizing and renaming the data were organized in a linear pattern and could be easily understood, added to, and redone at a later date if necessary. A lot of footwork was necessary to manually read through each survey and pull the appropriate questions. In addition to this, since each survey was unique in the arrangement that it presented its questions, in order to get a clear idea of which questions were available during which years, the corresponding question number for each year was indicated next to each qualifying question. Furthermore, there were occasions in which a question asked one year was given 4 possible answers, but the same question asked the next year was given 5 possible answers,. In cases such as this, reorganization of the possible question answers was necessary to maintain consistency among the years. Each of these steps was also outlined in the reference document in order to assist with future data use. The next step was inputting the variables into SAS® in a way that could be understood throughout the analysis. Renaming the variables with generic names (that were also matched within the reference document) was a necessary step for most of the variables; however, for variables in which the reorganization of answers was necessary, if/then/else statements were used. Here is an example of the coding used for inputting the data for the year 1991:

```
data YRBS1991; set mydata.YRBS1991;
      year=1991;
      AlcoholLife1=q33; AlcoholDay2=q34; AlcoholBinge4=q35; DemoAge1=q1;
      DemoSex2=q2; DemoEth3=q4; DrugsMarLife1=q37; DrugsMarDay2=q38;
      DrugsCocaLife1=q40; DrugsCocaDay2=q41; DrugsSteroLife1=q44;
      DrugsComboLife1=q43; EduRank2=q5; ExerHardActive1=q68; ExerStrength4=q70;
      ExerStretch5=q69; MoodConsiderS2=q19; MoodPlanS3=Q20; MoodAttemptS4=q21;
      MoodSeriousS5=q22; SexHist2=q48; SexAge3=q49; SexNumLife4=q50;
      SexNumMonth4=q51; SexSub5=q52; SexProtect7=q54; SexPregnant8=q55;
      SexSTD9=q56; TobacTry1=q23; TobacDaily2=q26; TobacDays4=q28;
      TobacAmount5=q29; VehicleHelmet1=q8; VehicleSelfSB2=q6;
      VehicleOtherD3=q11; VehicleSelfD3=q12; ViolMultWeap1=q14; ViolFight3=q16;
      ViolInjury3=q18; ViolWhom4=q17; WeightThink1=q57; WeightTry1=q58;

      if q59=1 or q59=5 then WeightDietExer2='2'; else if q59=2 or q59=3 or
      q59=4 then WeightDietExer2='1'; else WeightDietExer2=' ';
      if q60=1 or q60=5 then WeightPurge2='2'; else if q60=2 or q60=3 or q60=4
      then WeightPurge2='1'; else WeightPurge2=' ';
      if q31=1 then TobacChew7='2'; else if q31=2 or q31=3 or q31=4 then
      TobacChew7='1'; else TobacChew7=' ';
      if q30=2 then TobacQuit3='1'; else if q30=3 then TobacQuit3='2'; else
      TobacQuit3=' ';
      if q74=1 and q75=1 then ExerTeam6='1'; else if (q74=1 and q75=2) or
      (q74=2 and q75=1) then ExerTeam6='2'; else if (q74=1 and q75=3) or (q74=3
      and q75=1) or (q74=2 and q75=2) then ExerTeam6='3'; else if (q74=2 and
      q75=3) or (q74=3 and q75=2) or (q74=3 and q75=3) or q74=4 or q75=4 then
      ExerTeam6='4'; else ExerTeam6=' ';
      drop q1-q98 qn1-qn98 qnfrcig qnfrvg qnrovwgt qnovwgt qndlype qnanytob
      qnminpa qnnopa qnasatck bmipct ethorig raceorig qndepo qndepopl qndual
      qnfrvg2 qnfruit qnveg qnpa0day qnpa7day qnowt qnobese metrost qnstore
      qnabstsx qnrespsx qntencig q4orig;
      run;
```

Determination of Dependent Variables

In this study, the variable MoodConsider2 represents the question "During the past 12 months, did you ever seriously contemplate suicide?" This question was chosen for this study as a representation of suicidal ideation as an indicator of poor mental health. The above model was used to test correlations between suicidal thoughts and independent effects as well as interactions between the following variables: making a plan for suicide, number of times attempted suicide, and if severity of suicidal event lead to treatment by a medical professional. Each of these variables contain the values of each answer for their particular question. The results were outputted as Wald Chi-Square values.

The other variables chosen for this study as indicators of poor mental health were MoodDep1, WeightDietExer2, WeightFast2, WeightSupp2, WeightPurge2. MoodDep1 represented the question "During the past 12 months did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?". MoodDep1, therefore, was a question related to depression related symptoms as an indicator of poor mental health. All four weight variables could be seen as measures of disordered eating as indicators of poor mental health. WeightDietExer2 ("During the past 7 days, which one of the following did you do to lose weight or to keep from gaining weight?" [none, dieted, exercised]) and WeightPurge2 ("During the past 7 days, which one of the following did you do to lose weight or to keep from gaining weight?" [none, vomit, diet pills]). In addition to these two variables we then included WeightFast2("During the past 30 days, did you go without eating for 24 hours or more [also called fasting] to lose weight or to keep from gaining weight?") and WeightSupp2 ("During the past 30 days, did you take any diet pills, powders, liquids, without a doctor's advice to lose weight or to keep from gaining weight?").

Concatenating Data Sets

In order for data from all of the years to be used in this analysis, concatenating the 11 data sets was necessary. There is some debate in the research world as to whether combining separately administered data sets will affect the integrity of the results as the data collected were done so at considerably different times and by different people. For the purpose of this paper, we will consider that since each of surveys were administered under CDC guidelines to a specific target population in a controlled environment and supervised by professionals, that each of the years are similar enough in their methods of data collection, questionnaire length, and distribution of questions that we can be confident that any error created by the combining of these studies will be significant enough to affect the results. Please see WiiSU 2014 paper on concatenation versus separate analysis for more information on the impact of concatenation on survey analysis results (Schreiber-Gregory, 2014). The researcher for this study independently documented all of the questions used between the different surveys (a total of 168) and chose which variables would be used in the analysis based on the whether or not the questions fit either an aspect of mental health or risky behaviors. All questions asked between the years of 1991 and 2011 were included in the model and separated into categories based on risk behavior type or mental health concern. As mentioned above in the data cleaning section, the questions used were then given new names which made it simple for the appropriate questions to be concatenated together. As mentioned above, this was necessary because even though many of the questions used were present in all or most of the surveys, the order in which the questions appeared in the survey differed between each year.

The coding to concatenate the years together is given below:

```
data YRBS_Total;
set   YRBS1991   YRBS1993   YRBS1995   YRBS1997    YRBS1999    YRBS2001
      YRBS2003   YRBS2005   YRBS2007   YRBS2009    YRBS2011;
run;
```

Descriptive Statistics

To begin the analysis, the researcher used PROC SURVEYFREQ and PROC MEANS to get an idea of the data distribution and other descriptive statistics. Frequencies for demographics, risk behaviors, and mental health variables are all provided and reviewed. When viewing the output generated by these procedures, one must consider the current debate as to the appropriateness of weighting variables included in survey analyses of this size. Some research says that weighting variables is not error proof and can contribute to excluding important factors that would otherwise have shown significant in a nonweighted analysis. This could lead to losing some insight into significant contributing factors and therefore negatively affect the integrity and robustness of the model itself. Other research suggests that weighting the variables helps exclude variables with borderline significance that could muddy the significance and generalizability of the model. The appropriateness of weighting the variables involved in the model was explored using the results. An example of the code used is provided below:

```
proc means data=YRBS_Total;
run;

proc surveyfreq data=YRBS_Total;
      strata stratum;
      weight weight;
      by  AlcoholLife1 AlcoholDay2 AlcoholDaySP3 AlcoholBinge4 AlcoholGet5
      DrugsMarLife1 DrugsMarDay2 DrugsMarDaySP3 DrugsCocaLife1 DrugsCocaDay2
      DrugsInhaLife1 DrugsInhaDay2 DrugsHeroLife1 DrugsMethLife1
      DrugsSteroLife1 DrugsInjectLife1 DrugsEcstaLife1 DrugsPrescLife1
      DrugsComboLife1 ExerHardActive1 ExerHardActive2 ExerSoftActive1
      ExerStrength4 ExerStretch5 ExerTeam6 HealthDoctor1 HealthDentist2
      MoodDep1 MoodConsiderS2 MoodPlanS3 MoodAttemptS4 MoodSeriousS5 SexForce1
      SexHist2 SexAge3 SexNumLife4 SexNumMonth4 SexSub5 SexProtect7
      SexPregnant8 SexSTD9 TobacTry1 TobacDaily2 TobacQuit3 TobacDays4
```

```
    TobacDaysSP4 TobacAmount5 TobacGet6 TobacChew7 TobacChewSP8 TobacCigar9
    VehicleHelmet1 VehicleOtherSB2 VehicleSelfSB2 VehicleSelfSB2
    VehicleOtherD3 VehicleSelfD3 ViolMultWeap1 ViolMultWeapSP1 ViolGun1
    ViolUnsafe2 ViolThreatSP2 ViolDamageSP2 ViolFight3 ViolFightSP3
    ViolInjury3 ViolSigOth4 ViolWhom4 WeightTry1 WeightThink1 WeightDietExer2
    WeightFast2 WeightSupp2 WeightPurge2;
run;
```

The frequency statistics for the demographic variables are provided below. These statistics, when compared to the general population, are skewed enough to justify the need for weighting the variables. In order to get a more accurate picture of the generalizability of our final models to the population, the weights provided by the CDC were included in the factor and logistic regression analyses used later in this study.

Table of DemoAge1

DemoAge1	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
1	215	159.98542	16.06412	0.1004	0.0101
2	175	166.16264	18.69376	0.1043	0.0117
3	14694	16739	194.56691	10.5054	0.1191
4	35410	38904	280.41672	24.4163	0.1661
5	40895	41623	287.38032	26.1225	0.1696
6	41463	38958	272.74984	24.4498	0.1639
7	26424	22788	218.55120	14.3014	0.1338
Total	159276	159338	361.05816	100.000	
Frequency Missing = 415					

Table 1: This table displays frequency distributions for the age demographics. Variable values are as follows: 1 - 12 years old or younger, 2 - 13 years old, 3 - 14 years old, 4 - 15 years old, 5 - 16 years old, 6 - 17 years old, 7 - 18 years old or older

Table of DemoSex2

DemoSex2	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
1	80751	77374	344.89406	48.5748	0.1921
2	78439	81914	365.97221	51.4252	0.1921
Total	159190	159287	361.02033	100.000	
Frequency Missing = 501					

Table 2: This table displays frequency distributions for the gender demographics. Variable values are as follows: 1 - Female, 2 – Male

Table of DemoEth3

DemoEth3	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
1	65428	100120	363.07922	63.0682	0.1651
2	37028	22170	168.15640	13.9655	0.1055
3	40162	19666	138.79944	12.3882	0.0895
4	6228	6068	102.26670	3.8224	0.0645
5	1862	1316	46.08555	0.8293	0.0291
6	8012	9408	186.20229	5.9265	0.1136
Total	158720	158749	359.47341	100.000	
Frequency Missing = 971					

Table 3: This table displays frequency distributions for the race/ethnicity demographics. Variabe values are as follows: 1 - White, 2 - Black or African American, 3 - Hispanic or Latino, 4 - Asian or Pacific Islander, 5 - Native American or Alaskan Native, 6 - Other

The second goal of calculating these frequency statistics was to explore the distribution of response to risk behavior and mental health variables. These frequencies showed very little change in each of the responses over the years. Also, when looking at the percentages of each response, the majority of students either denied participating in any unique risky behavior or reported participating in the behavior at a lower rate than other respondents. Given these results, the researcher sought to find out if participation in a particular set of risky behaviors, being that any unique risky behavior is avoided by the majority of the population, would contribute to suicidal ideation. This idea was formulated from the general idea that most risky behaviors are viewed as poor decisions or compensatory behaviors initiated by the environment or other stimuli.

Demographic Results

Demographic differences in risk behaviors and mental health issues are explored below. The demographic variables explored in this analysis are biological age (regardless of grade), gender, and race/ethnicity. Demographic variables of race and ethnicity were included in this analysis but separately recoded by the researcher on account of the high variability in answer options between the survey years. Average amount of sleep per night is also explored in this section as an educational question as to the effect of sleep on mental health. Significant results from this analysis are discussed and will be included in the final model as an additional component to the interactions between risk behaviors and mental health issues.

Logistic Regression

A logistic analysis was conducted in order to test how much of the variability in mental health responses could be explained by demographic variables. The logistic analysis was written in a manner so that a multiple regression analysis could be performed, given that the particular variables used were categorical. Also, given that the variables used are in a complex survey format, PROC SURVEYLOGISTIC was a necessary procedure to employ for this analysis as it accounts for complex survey designs.

```
proc surveylogistic data=YRBS Total;
      class MoodConsider2 DemoAge1 DemoSex2 DemoEth3;
      cluster psu;
      strata stratum;
      model MoodConsider2 = DemoAge1 DemoSex2 DemoEth3 / rsq;
      weight weight;
run;
```

Nesting options were used within PROC SURVEYLOGISTIC throughout this paper in order to account for the combination of the different data sets and to account for the structure used within these data sets. Since the data sets used are for the same region (national) and differ mainly in the years that they were given, the nesting options available remain consistent across the years and do not need to be readjusted. The nesting option `cluster` was used in order to account for survey degrees of freedom. According to the CDC (2014), SAS® considers survey degrees of freedom to be the difference between the number of PSUs and the number of first stage sampling strata among strata and PSUs that contain at least one observation with a value for the variable(s) within the analysis. Considering the possible occurrence in subpopulation analysis (an occurrence that is potentially increased in probability with multiple datasets) that an analytic variable is missing for all survey respondents in one or more PSU or stratum, then an alternate definition of survey degrees of freedom is needed in order to avoid overestimation. This alternative definition has in fact been defined and recommended by Korn and Graubard (1999) and is used by the CDC as the PSU variable. Therefore, in defining `cluster psu`, the programmer is enabling SAS® to correctly calculate the degrees of freedom without the threat of overestimation. As an added bonus to this option, the log will indicate if there were empty clusters for a variable and how many of the clusters available were included in the analysis. The `strata` statement is also used in order to indicate the name of the stratification variable needed for this study (stratum). Lastly, the `weight` option is used in order to indicate the name of the weight variable to be used in the analysis (weight). All of these nested options assist with ensuring an appropriately coded and rounded data set for analysis.

It is also useful to note that max-rescaled r-square estimates were used in this analysis to approximate model fit. However, there is some thought as to the relevance of using this estimate, which an available option through the SURVEYLOGISTIC procedure in SAS® by indicating `rsq` at the end of a forward slash in model statement, as it is not viewed as the most accurate estimate of model fit. Alternatives to this model were presented in a SAS® Global Forum 2014 paper by Paul D. Allison and are noted in the references of this paper. These same alternatives are also touched on in the accompanying WiilSU 2014 paper by Schreiber-Gregory in relation to this data set.

Depression

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of depression. Max rescaled R-square values indicated that this demographic model only contributes to about 5% of the variability in depression. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, adolescents in high school that are twelve years or younger are twice as likely than their older peers to identify feelings of depression. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were twice as likely than males to identify feelings of depression. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, Latinos, and American Indians/Alaskan

Natives were significantly more likely to identify feelings of depression than their peers while Whites and African Americans were significantly less likely to identify these same feelings.

When analyzing sleep, our model indicated that less sleep did significantly contribute to feelings of depression students who indicated getting more sleep were significantly less likely to identify these feelings. Max rescaled R-square values indicated that the model of sleep only contributed to about 5% of the variability in depression.

Suicidal Ideation

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of suicidal ideation. Max rescaled R-square values indicated that this demographic model only contributes to about 2.5% of the variability in suicidal ideation. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, adolescents in high school that are twelve years or younger are five times as likely than their older peers to identify feelings of suicidal ideation. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were almost twice as likely than males to identify feelings of suicidal ideation. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, American Indians/Alaskan Natives were significantly more likely to identify feelings of suicidal ideation than their peers while Whites and African Americans were significantly less likely to identify these same feelings.

When analyzing sleep, our model indicated that less sleep did significantly contribute to feelings of suicidal ideation while students who identified with getting more sleep were significantly less likely to identify these feelings. Max rescaled R-square values indicated that the model of sleep only contributed to about 5% of the variability in suicidal ideation.

Attempt Suicide

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of suicide attempts. Max rescaled R-square values indicated that this demographic model only contributes to about 8% of the variability in suicide attempts. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, adolescents in high school that are twelve years or younger were significantly less likely than their older peers to identify having attempted suicide. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were significantly less likely than males to identify having attempted suicide. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, Latinos, and American Indians/Alaskan Natives were significantly less likely to identify having attempted suicide than their peers while Whites and African Americans and Asian Americans/Pacific Islanders were significantly more likely to identify these same actions.

When analyzing sleep, our model indicated that less sleep did significantly contribute to actions of attempted suicide while students who indicated getting more sleep were significantly less likely to identify these actions. Max rescaled R-square values indicated that the model of sleep only contributed to about 6% of the variability in suicide attempts.

Diet and Exercise

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of Diet/Exercise. Max rescaled R-square values indicated that this demographic model only contributes to about 5% of the variability in Diet/Exercise. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, there was no significant contribution to this model. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were significantly more likely than males to identify using fasting to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, Asian Americans and Pacific Islanders were significantly more likely to diet and exercise to lose weight than their peers while African Americans were significantly less likely to diet and exercise to lose weight.

When analyzing sleep, our model indicated that sleep did significantly contribute to using fasting as a way to lose weight. Max rescaled R-square values indicated that the model of sleep only contributed to about .2% of the variability in using diet and exercise as a way to lose weight.

Fasting

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of fasting. Max rescaled R-square values indicated that this demographic model only contributes to about 4.8% of the variability in fasting. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, students who were 12 years of age or younger were 60% more likely to fast to lose weight than older peers. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were significantly more likely than males to identify using fasting to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, Asian Americans/Pacific Islanders and Native American/Alaskan Natives were significantly more likely to fast to lose weight than their peers while Whites and African Americans were significantly less likely to fast to lose weight.

When analyzing sleep, our model indicated that sleep did significantly contribute to using fasting as a way to lose weight. Students who indicated less sleep were more likely to identify having fasted than students how had more sleep. Max rescaled R-square values indicated that the model of sleep only contributed to about 4.8% of the variability in using fasting as a way to lose weight.

Diet Supplements

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of Diet Supplement usage. Max rescaled R-square values indicated that this demographic model only contributes to about 2.3% of the variability in Diet Supplement usage. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, students who were 12 years of age or younger were about 8 times more likely to use diet supplements to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were significantly more likely than males to identify using diet supplements to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, African Americans were significantly less likely to use diet supplements to lose weight than their peers

When analyzing sleep, our model indicated that sleep did significantly contribute to using diet supplements as a way to lose weight. Max rescaled R-square values indicated that the model of sleep only contributed to about 2.8% of the variability in using diet supplements as a way to lose weight.

Purging

Demographic variables of age, gender, and ethnicity all contributed significantly to the demographic model of Purging. Max rescaled R-square values indicated that this demographic model only contributes to about 4.2% of the variability in Purging. Even so, it is useful to consider demographics as biological and environmental contributing factors to this mental illness variable. The analysis of maximum likelihood estimates and odds ratio scores indicated that for age, students who were 12 years of age or younger were almost 15 times more likely than their peers to identify with purging to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for gender, females were almost three times as likely than males to identify using purging to lose weight. The analysis of maximum likelihood estimates and odds ratio scores indicated that for ethnicity/race, Native American/Alaskan Natives were significantly more likely to purge to lose weight than their peers while African Americans and Whites were significantly less likely to purge to lose weight.

When analyzing sleep, our model indicated that sleep did significantly contribute to using purging as a way to lose weight. Max rescaled R-square values indicated that the model of sleep only contributed to about 4.5% of the variability in using purging as a way to lose weight.

Factor Analysis

A factor analysis was performed next in order to test the correlations between the different variables and to check for underlying dimensions of related variables (Child, 1990). The variables chosen for each factor analysis were chosen based on their base similarities such as alcohol use, drug use, sexuality, health care, weight, mood, tobacco use, violence, and vehicle safety. Since this factor analysis used character data in both binary and Likert scale formats, a limitation arises. Since binary data contains only two data points and Likert scale data contains more than two points, magnitudes of correlations between these variables shrink due to the range restriction. In order to control for this a polychoric correlation matrix was needed. SAS® provides such a matrix in a macro available http://support.sas.com/kb/25/addl/fusion25010_1_polychor.sas.txt. The polychoric correlation matrix from SAS® can be implemented in two steps: (1) by first initializing the macro and computing the polychoric correlation matrix and (2) submitting the computed matrix to PROC FACTOR for factor extraction. An example of the coding is provided below:

```
data YRBS Total FA;
      set   YRBS1991     YRBS1993     YRBS1995     YRBS1997     YRBS1999
            YRBS2001     YRBS2003     YRBS2005     YRBS2007     YRBS2009
      YRBS2011;

run;
```

```
%polychor(data=YRBS_Total_FA, var=AlcoholLife1 AlcoholDay2 AlcoholDaySP3
AlcoholBinge4 AlcoholGet5 DrugsMarLife1 DrugsMarDay2 DrugsMarDaySP3
DrugsCocaLife1 DrugsCocaDay2 DrugsInhaLife1 DrugsInhaDay2 DrugsHeroLife1
DrugsMethLife1 DrugsSteroLife1 DrugsInjectLife1 DrugsEcstaLife1 DrugsPrescLife1
DrugsComboLife1 ExerHardActive1 ExerHardActive2 ExerSoftActive1 ExerStrength4
ExerStretch5 ExerTeam6 HealthDoctor1 HealthDentist2 MoodDep1 MoodConsiderS2
MoodPlanS3 MoodAttemptS4 MoodSeriousS5 SexForce1 SexHist2 SexAge3 SexNumLife4
SexNumMonth4 SexSub5 SexProtect7 SexPregnant8 SexSTD9 TobacTry1 TobacDaily2
TobacQuit3 TobacDays4 TobacDaysSP4 TobacAmount5 TobacGet6 TobacChew7
TobacChewSP8 TobacCigar9 VehicleHelmet1 VehicleOtherSB2 VehicleSelfSB2
VehicleSelfSB2 VehicleOtherD3 VehicleSelfD3 ViolMultWeap1 ViolMultWeapSP1
ViolGun1 ViolUnsafe2 ViolThreatSP2 ViolDamageSP2 ViolFight3 ViolFightSP3
ViolInjury3 ViolSigOth4 ViolWhom4 WeightTry1 WeightThink1 WeightDietExer2
WeightFast2 WeightSupp2 WeightPurge2,out=YRBS_Total, type=corr);

proc corr data=YRBS_Total nocorr alpha nomiss;
      var AlcoholLife1 AlcoholDay2 AlcoholDaySP3 AlcoholBinge4 AlcoholGet5;
run;

proc factor data=YRBS_Total
      method=prinit
      priors=smc
      scree
      residuals
      rotate=promax
      corr
      heywood;
      var AlcoholLife1 AlcoholDay2 AlcoholDaySP3 AlcoholBinge4 AlcoholGet5;
run;
```

As seen in this sample code, proc factor for the alcohol variable w as invoked using `method=prinit`, `priors=smc`, `scree`, `residuals`, `rotate=promax`, `corr`, and `heywood`. The option `method=prinit` requests that an iterated principal factor analysis be used. The option `priors=smc` requests that squared multiple correlations betw een a given input variable and the other variables in the model be used to estimate the variable's prior communality. The option `scree` requests that a scree plot of the eigenvalues be displayed in the output. The option `corr` requests that both a correlation matrix and partial correlation matrix be displayed in the output. The option `residuals` requests that a residual correlation matrix and associated partial correlation matrix be displayed for the factor analysis in the output as w ell. The option `rotate=promax` requests that an orthogonal promax rotation be performed on the resulting factors. This w as chosen based on the fact that after the initial factor extraction, orthogonal transformation, and varimax transformation, common factors w ere found to remain uncorrelated w ith each other and therefore required a promax rotation to ensure that a given variable w ould only have a high loading on one factor and a near zero loading on other factors. Given the diversity and complexity of the data used, this w as a necessary request. Finally, the option `Heywood` requests that any communality greater than 1, be set to 1, allow ing iterations to proceed.

Some minor adjustments to the factor analysis for the different variable groups w ere needed based on an individual basis. For factor groups Drug Use, Tobacco Use, and Violence, `priors` was set to `max` instead of `smc` based on the need for the prior communality estimate for each of the variables w ithin these groups to be set to its maximum absolute correlation w ith any other variable. For the factor group Vehicle Safety, `maxiter` was set to `100` based on this variable groups need to limit the maximum number of iterations for factor extraction, as it w as exceeded w hen using the default of 30.

Results of these factor analyses are provided below. Please see the reference Microsoft Excel® document attached to these proceedings for the explanation and distribution of the questions represented by each of the variables mentioned. Variables that fell into the categories of sleep, education, and demographics w ere not included in any of the factor analyses as there w as either little chance that they w ould correlate or distribute in a meaningful manner w ith other variables (ex: gender and ethnicity should be evenly distributed w ithin each other) or they w ere available in a very limited number of years.

Alcohol Use: For alcohol use this study looked at variables: AlcoholLife1, AlcoholDay2 AlcoholDaySP3, AlcoholBinge4, and AlcoholGet5. Three factors w ere indentified for these variables. When considering factor loading, the first factor contained high loading for variables AlcoholLife1, AlcoholDay2, and AlcoholDaySP3. The second factor contained high loading for variable AlcoholBinge4, and the fourth factor contained high loadings for variables AlcoholGet5. Variables chosen for inclusion in the logistic regression analysis w ere then reduced to recent uses of alcohol and the other tw o identified factors: AlcoholDay2 and AlcoholBinge4.

Drug Use: For drug use this study looked at variables: DrugsMarLife1, DrugsMarDay2, DrugsMarDaySP3, DrugsCocaLife1, DrugsCocaDay2, DrugsInhaLife1, DrugsInhaDay2, DrugsHeroLife1, DrugsMethLife1, DrugsSteroLife1, DrugsInjectLife1, DrugsEcstaLife1, DrugsPrescLife1, and DrugsComboLife1. Twelve different factors were identified for these variables. When considering factor loading, the first factor contained high loading for variables DrugsMarLife1, DrugsMarDay2, DrugsMarDaySP3. The second factor contained high loading for variables DrugsInhaLife1, DrugsInhaDay2. The third factor contained high loading for variables DrugsCocaLife1, DrugsCocaDay2. The fourth factor contained high loading for variable DrugsSteroLife1. The fifth factor contained high loading for variables DrugsComboLife1. The sixth factor contained high loading for variable DrugsEcstaLife1. The seventh factor contained high loading for variable DrugsMethLife1. The eighth factor contained high loading for variable DrugInjectLife1. The eleventh factor contained high loading for variable DrugsHeroLife1. The ninth, tenth, and twelfth factors contained varying loadings of the different variables mentioned. Variables chosen for inclusion in the logistic regression analysis were then reduced to recent uses of drugs and the other identified factors: DrugsMarDay2, DrugsInhaDay2, DrugsCocaDay2, DrugsSteroLife1, DrugsComboLife1, DrugsEcstaLife1, DrugsMethLife1, DrugsInjectLife1, and DrugsHeroLife1.

Exercise Participation: For exercise participation this study looked at variables: ExerHardActive1, ExerSoftActive1, ExerStrength4, ExerStretch5, ExerTeam6, and ExerInjury7. Only one factor was identified for these variables. Therefore, the variable of ExerHardActive1 was chosen to be included in the regression analysis based on being the highest loading variable for that factor.

Mental Health: For mental health variables this study looked at variables: MoodDep1, MoodConsiderS2, MoodPlanS3, MoodAttemptS4, and MoodSeriousS5. Two factors were identified for these variables. The first factor identified had high loadings for MoodDep1, MoodConsiderS2, and MoodPlanS3. The second factor identified had high loadings for MoodAttemptS4 and MoodSeriousS5. Therefore, the variables chosen to be included in the regression analysis based on their high loadings were MoodDep1, MoodConsiderS2, and MoodAttemptS4.

Sexuality: For sexuality variables this study looked at variables: SexForce1, SexHist2, SexAge3, SexNumLife4, SexNumMonth4, SexSub5, SexProtect7, SexPregnant8, and SexSTD9. Only one factor was identified for these variables. Therefore SexForce1, SexHist2, SexNumMonth4, SexPregnant8, and SexSTD9 were identified and will be included in the regression analysis based on their relevance and high factor loadings.

Tobacco Use: For tobacco use this study looked at variables: TobacTry1, TobacDaily2, TobacQuit3, TobacDays4, TobacDaysSP4, TobacAmount5, TobacGet6, TobacChew7, TobacChewSP8, and TobacCigar9. Three factors were identified for these variables. The first factor identified included variables TobacTry1, TobacDaily2, TobacQuit3, TobacDays4, TobacDaysSP4, and TobacAmount5. The second factor identified included variables TobacChewSP8, TobacChew7, and TobacCigar9. The final variable identified included TobacGet6. Based on factor loadings and considering that mental health issues were measured on a recent basis, variables TobacDays4, TobacTry1, and TobacChew7 will be included in the regression analysis.

Vehicle Safety: For vehicle safety this study looked at variables: VehicleHelmet1, VehicleOtherSB2, VehicleSelfSB2, VehicleOtherD3, and VehicleSelfD3. Two factors were identified for these variables. The first factor included high loadings for VehicleOtherD3, and VehicleSelfD3. The second variable included high factor loadings for VehicleOtherSB2. Considering the relevance of each of these identified variables, VehicleOtherD3, VehicleSelfD3, and VehicleOtherSB2 will all be included in the regression analysis.

Violence Exposure/Participation: For exposure or participation in violence this study looked at variables: ViolMulWeap1, ViolMultWeapSP1, ViolGun1; ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, ViolFight3, ViolFightSP3, ViolInjury3, ViolSigOth4, and ViolWhom4. A total of nine factors were identified for these variables. The first factor contained high loading for variables ViolMultWeap1, ViolMultWeapSP1, ViolGun1 The second factor contained high loading for variables ViolFightSP3 The fourth factor contained high loading for variables ViolUnsafe2, ViolFight3 The fifth factor contained high loading for variables ViolFight3, ViolSigOth4 The sixth factor contained high loading for variables ViolInjury3 The seventh factor contained high loading for variables ViolThreatSP2 The eighth factor contained high loading for variables ViolDamageSP2 The ninth and third factor contained varying loadings of multiple variables. No high loadings for any one variable were present in these last two factors. Considering the above, variables ViolMultWeap1, ViolFightSP3, ViolFight3, ViolUnsafe2, ViolSigOth4, ViolInjury3, ViolThreatSP2, and ViolDamageSP2 will all be included in the regression analysis.

Weight: For weight this study looked at variables: WeightThink1, WeightTry1, WeightDietExer2, WeightFast2, WeightSupp2, and WeightPurge2. Two factors were identified for these variables. The first factor contained high loadings for variables WeightFast2, WeightSupp2, and WeightPurge2. The second variable contained high loadings for variable WeightDietExer2 and WeightTry1. Variable WeightThink1 did not have high loadings in either factor. Considering these results, variables WeightDietExer2, WeightFast2, WeightSupp2, and WeightPurge2 will all be included in the regression analysis.

Please contact the author with any questions on this section. Additional information on latent factors and specific statistics are available upon request.

Final Models

The main interest of the final model results are the presence of significant correlations between the mental health variables and different health-risk behaviors. Variables that displayed high factor loadings for separate factors within an identified risk-behavior cohort are included in this model. A sample coding for these variables is provided below:

```
proc surveylogistic data=YRBS Total;
      class MoodDep1 AlcoholDay2 DrugsInhaDay2 DrugsMethLife1 DrugsEcstaLife1
      ExerHardActive1 SexForce1 TobacTry1 TobacDays4 TobacChew7 ViolMultWeap1
      ViolUnsafe2 ViolDamageSP2 ViolFight3 ViolSigOth4;
      cluster psu;
      strata stratum;
      model MoodDep1 = AlcoholDay2 DrugsInhaDay2 DrugsMethLife1 DrugsEcstaLife1
      ExerHardActive1 SexForce1 TobacTry1 TobacDays4 TobacChew7 ViolMultWeap1
      ViolUnsafe2 ViolDamageSP2 ViolFight3 ViolSigOth4 / rsq;
      weight weight;
run;
```

Depression

The final model of depression supports that the following variables contribute significantly to the variability in responses to the question MoodDep1: AlcoholDay2, DrugsInhaDay2, DrugsMethLife1, DrugsEcstaLife1, ExerHardActive1, SexForce1, TobacTry1, TobacDays4, TobacChew7, ViolMultWeap1, ViolUnsafe2, ViolDamageSP2, ViolFight3, and ViolSigOth4. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of depressive symptoms. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	13478.097	11930.909
SC	13485.454	12379.672
-2 Log L	13476.097	11808.909

-Square	0.1339	Max-rescaled R-Square	0.1947

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1657.7514	60	<.0001
Score	1683.7315	60	<.0001
Wald	9738.9127	60	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AlcoholDay2	6	33.7364	<.0001
DrugsInhaDay2	5	44.2964	<.0001
DrugsMethLife1	5	13.6903	0.0177
DrugsEcstaLife1	5	23.6901	0.0002
ExerHardActive1	7	40.2238	<.0001
SexForce1	1	50.4077	<.0001
TobacTry1	1	19.3990	<.0001
TobacDays4	6	15.4489	0.0170
TobacChew7	1	15.7040	<.0001
ViolMultWeap1	4	18.9681	0.0008
ViolUnsafe2	4	99.2497	<.0001
ViolDamageSP2	7	87.7285	<.0001
ViolFight3	7	75.0762	<.0001
ViolSigOth4	1	42.4344	<.0001

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 20% of the variability in depression identification. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsInhaDay2, DrugsMethLife1, DrugsEcstaLife1, TobacDays4, ViolMultWeap1, ViolUnsafe2, ViolDamageSP2, and ViolFight3 were significantly less likely to identify feelings of depression than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying these feelings while students who indicated greater participation were more likely to identify these feelings regardless of number of times of usage/participation. On the other hand, students who identified less participation in challenging physical activities were significantly more likely to identify feelings of depression than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. ViolSigOth4 indicated that students who indicated having experienced violence from a significant other were almost 2 times more likely to identify feelings of depression than peers who denied this variable. For TobacChew7, students who identified participation chewing tobacco were significantly more likely to also identify feelings of depression than their peers who denied participation in this variable. Variable TobacTry1 indicated that students who tried cigarettes were 1.5 times more likely to identify depression symptoms while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were 2.6 times more likely to identify these same feelings. This model suggests that when looking at depression as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Suicidal Ideation

The final model of suicidal ideation supports that the following variables contribute significantly to the variability in responses to the question MoodConsiderS2: AlcoholDay2, DrugsInhaDay2, DrugsMethLife1, DrugsSteroLife1, ExerHardActive1, SexForce1, TobacDays4, TobacChew7, VehicleSelfD3, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, and ViolSigOth4. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of suicidal ideation. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	9498.591	8432.139
SC	9505.903	8870.875
-2 Log L	9496.591	8312.139

R-Square 0.1014	Max-rescaled R-Square 0.1762

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1053.4917	71	<.0001
Score	1124.5604	71	<.0001
Wald	1638364012	71	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AlcoholDay2	6	36.2926	<.0001
DrugsInhaDay2	5	65.9721	<.0001
DrugsMethLife1	5	53.6327	<.0001
DrugsSteroLife1	5	28.9721	<.0001
ExerHardActive1	7	25.6334	0.0006
SexForce1	1	87.5406	<.0001
TobacDays4	6	38.9449	<.0001
TobacChew7	1	9.1733	0.0025
VehicleSelfD3	4	10.9159	0.0275
ViolUnsafe2	4	21.7489	0.0002
ViolThreatSP2	7	17.1693	0.0163
ViolDamageSP2	7	21.8359	0.0027
ViolSigOth4	1	17.9589	<.0001

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 17.6% of the variability in suicidal ideation identification. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsInhaDay2, DrugsMethLife1, DrugsSteroLife1, TobacDays4, VehicleSelfD3, ViolUnsafe2, ViolThreatSP2, and ViolDamageSP2 were significantly less likely to identify thoughts of suicide than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying these thoughts while students who indicated greater participation were more likely to identify these thoughts regardless of number of times of usage/participation. On the other hand, students who identified less participation in challenging physical activities were significantly more likely to identify suicidal thoughts than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. ViolSigOth4 indicated that students who indicated having experienced violence from a significant other were almost 64% more likely to identify suicidal thoughts than peers who denied this variable. For TobacChew7, students who identified participation chewing tobacco were significantly more likely to also identify suicidal thoughts than their peers who denied participation in this variable while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were 2.5 times more likely to identify these same thoughts. As with depression, this model suggests that when looking at suicidal ideation as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Attempt Suicide

The final model of suicide attempts supports that the following variables contribute significantly to the variability in responses to the question MoodAttemptS2: AlcoholDay2, DrugsInhaDay2, DrugsHeroLife1, DrugsMethLife1, DrugsSteroLife1, ExerHardActive1, SexForce1, TobacTry1, TobacChew7, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, ViolFightSP3, and ViolSigOth4. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of recent suicide attempts. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	7104.903	5822.002
SC	7133.719	6326.269
-2 Log L	7096.903	5682.002

R-Square 0.1327	Max-rescaled R-Square 0.2600

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	1414.9014	66	<.0001
Score	2130.0006	66	<.0001
Wald	.	65	.

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
AlcoholDay2	6	19.3398	0.0036
DrugsInhaDay2	5	79.4713	<.0001
DrugsHeroLife1	5	23.3462	0.0003
DrugsMethLife1	5	36.3919	<.0001
DrugsSteroLife1	5	21.5946	0.0006
ExerHardActive1	7	38.0853	<.0001
SexForce1	1	113.8444	<.0001
TobacTry1	1	6.6820	0.0097
TobacChew7	1	20.3622	<.0001
ViolUnsafe2	4	46.9804	<.0001
ViolThreatSP2	7	34.1676	<.0001
ViolDamageSP2	7	65.3249	<.0001
ViolFightSP3	7	177.5436	<.0001
ViolSigOth4	1	10.5652	0.0012

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 26% of the variability in identification of recent suicide attempts. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsInhaDay2, DrugsHeroLife1, DrugsMethLife1, DrugsSteroLife1, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, and ViolFightSP3 were significantly more likely to deny recent suicide attempts than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying having ever attempted suicide while students who indicated greater participation were more likely to identify these attempts regardless of number of times of usage/participation. On the other hand, students who identified less participation in challenging physical activities were significantly more likely to identify recent suicide attempts than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. ViolSigOth4 indicated that students who indicated having experienced violence from a significant other were significantly more likely to identify having attempted suicide than peers who denied this variable. For TobacChew7, students who identified participation chewing tobacco were significantly more likely to also identify having attempted suicide than their peers who denied participation in this variable while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were also significantly more likely to identify having attempted suicide recently. As with depression and suicidal ideation, this model suggests that when looking at suicide attempts as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Diet and Exercise

The final model of weight loss by diet/exercise supports that the following variables contribute significantly to the variability in responses to the question WeightDietExer2: AlcoholDay2, DrugsMarDay2, DrugsMethLife1, DrugsSteroLife1, DrugsComboLife1, ExerHardActive1, SexForce1, SexNumMonth4, TobacChew7, VehicleSelfD3, ViolMultWeap1, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, ViolInjury3, and ViolSigOth4. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of using diet and exercise to lose weight. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	12927.271	12425.375
SC	12934.534	12962.811
-2 Log L	12925.271	12277.375

R-Square	0.0596	Max-rescaled R-Square	0.0844

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	647.8961	73	<.0001
Score	633.6052	73	<.0001
Wald	4786840.92	73	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AlcoholDay2	6	39.7751	<.0001
DrugsMarDay2	5	11.3923	0.0441
DrugsMethLife1	5	16.0261	0.0068
DrugsSteroLife1	5	10.9916	0.0515
DrugsComboLife1	5	49.6549	<.0001
ExerHardActive1	7	188.7903	<.0001
SexForce1	1	39.4401	<.0001
SexNumMonth4	7	30.3844	<.0001
TobacChew7	1	9.3435	0.0022
VehicleSelfD3	4	12.9705	0.0114
ViolMultWeap1	4	19.8341	0.0005
ViolUnsafe2	4	9.5603	0.0485
ViolThreatSP2	7	477.1294	<.0001
ViolDamageSP2	7	26.4335	0.0004
ViolInjury3	4	274.6227	<.0001
ViolSigOth4	1	5.3254	0.0210

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 8% of the variability in identification of using diet and exercise to lose weight. Analysis of Maximum Likelihood

Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsMarDay2, DrugsMethLife1, DrugsSteroLife1, DrugsComboLife1, SexNumMonth4, VehicleSelfD3, ViolMultWeap1, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, and ViolInjury3 were significantly more likely to identify using diet and exercise to lose weight than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of identifying that they used diet and exercise to lose weight while students who indicated greater participation were less likely to identify this type of weight loss regardless of number of times of usage/participation. On the other hand, students who identified less participation in challenging physical activities were significantly less likely to identify using diet and exercise for weight loss than peers who identified greater participation in these activities. Similar results were found for binary variables. ViolSigOth4 indicated that students who indicated having experienced violence from a significant other were significantly more likely to identify using diet and exercise to lose weight than peers who denied this variable. For TobacChew7, students who identified participation in chewing tobacco were significantly less likely to also identify using diet and exercise to lose weight than their peers who denied participation in this variable while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were 2 times more likely to identify using diet and exercise to lose weight. The differing results between this mental health variable and the previous three could be summed up to the fact that diet and exercise are generally considered a healthy way to lose weight. It can be used in unhealthy ways and an unhealthy use of diet and exercise should be watched closely by clinicians and researchers. However, this model supports that diet and exercise seem to be the go to method of weight loss for individuals who have less or no engagement in other risky behaviors.

Fasting

The final model of using fasting for weight loss supports that the following variables contribute significantly to the variability in responses to the question WeightFast2: AlcoholDay2, DrugsInhaDay2, DrugsHeroLife1, DrugsMethLife1, DrugsComboLife1, ExerHardActive1, SexForce1, ViolDamageSP2, ViolFightSP3, and ViolSigOth4. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of using fasting for weight loss. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	8194.462	7595.375
SC	8201.814	7962.963
-2 Log L	8192.462	7495.375

R-Square	0.0587	Max-rescaled R-Square	0.1154

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	697.0867	49	<.0001
Score	885.6637	49	<.0001
Wald	72157138.0	49	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AlcoholDay2	6	29.2804	<.0001
DrugsInhaDay2	5	39.5685	<.0001
DrugsHeroLife1	5	15.1950	0.0096
DrugsMethLife1	5	35.4304	<.0001
DrugsComboLife1	5	24.5951	0.0002
ExerHardActive1	7	29.4870	0.0001
SexForce1	1	34.5465	<.0001
ViolDamageSP2	7	27.7930	0.0002
ViolFightSP3	7	126.4956	<.0001
ViolSigOth4	1	7.8441	0.0051

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 11.5% of the variability in identification of using fasting for weight loss. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsInhaDay2, DrugsHeroLife1, DrugsMethLife1, DrugsComboLife1, and ViolDamageSP2 were significantly more likely to use fasting to lose weight than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying having ever attempted suicide while students who indicated greater participation were more likely to identify using fasting for weight loss regardless of number of times of usage/participation. On the other hand, students who identified less participation in challenging physical activities (ExerHardActive1) or fights (ViolFight3) were significantly more likely to identify using fasting for weight loss than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. ViolSigOth4 indicated that students who indicated having experienced violence from a significant other were1.5 times more likely to identify having used fasting for weight loss than peers who denied this variable while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were also 2.6 times more likely to identify having recently used fasting for weight loss. As with some of the previous mental health variables, this model suggests that when looking at fasting (defined as an unhealthy method of weight loss) as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Diet Supplements

The final model of using diet supplements without a doctors order to lose weight supports that the following variables contribute significantly to the variability in responses to the question WeightSupp2: AlcoholDay2, DrugsMethLife1, DrugsSteroLife1, SexForce1, SexHist2, SexNumMonth4, TobacTry1, TobacDays4, VehicleOtherSB2, VehicleSelfD3, ViolThreatSP2, and ViolDamageSP2. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of diet supplement use. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	5487.172	4859.750
SC	5494.449	5259.983
-2 Log L	5485.172	4749.750

R-Square 0.0665	Max-rescaled R-Square 0.1656

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	735.4220	54	<.0001
Score	969.4442	54	<.0001
Wald	1.27534E13	54	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
AlcoholDay2	6	55.6641	<.0001
DrugsMethLife1	5	50.4493	<.0001
DrugsSteroLife1	5	26.5424	<.0001
SexForce1	1	18.0237	<.0001
SexHist2	1	7.5174	0.0061
SexNumMonth4	6	14.0816	0.0287
TobacTry1	1	10.4908	0.0012
TobacDays4	6	26.0576	0.0002
VehicleOtherSB2	5	13.7608	0.0172
VehicleSelfD3	4	21.4363	0.0003
ViolThreatSP2	7	16.6676	0.0197
ViolDamageSP2	7	40.0729	<.0001

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 16.6% of the variability in identification of recent diet supplement use. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in AlcoholDay2, DrugsMethLife1, DrugsSteroLife1, SexNumMonth4, TobacDays4, and VehicleSelfD3, were significantly more likely to deny recent diet supplement use than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying having ever used diet supplements while students who indicated greater participation were more likely to identify these attempts regardless of number of times of usage/participation. On the other hand, students who identified less participation in VehicleOtherSB2, ViolThreatSP2, and ViolDamageSP2 were significantly more likely to identify recent diet supplement use than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. SexHist2 supported that students who denied having sex in the past were significantly less likely to identify having used diet supplements to lose weight than peers who admitted to this variable while SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were almost 2 times more likely to identify having used diet supplements recently. For TobacTry7 however, students who identified having never tried smoking cigarettes in the past were almost 2 times more likely to use diet supplements than other peers. As with most other mental health variables in this study, this model suggests that when looking at unprescribed diet supplement use as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Purging

The final model of purging to lose weight supports that the following variables contribute significantly to the variability in responses to the question WeightPurge2: DrugsCocaDay2, DrugsInhaDay2, DrugsMethLife1, SexForce1, TobacDays4, VehicleOtherD3, VehicleSelfD3, ViolUnsafe2, ViolThreatSP2, ViolDamageSP2, ViolFightSP3, and ViolInjury3. More specifically, this model suggests that unhealthy participation in these activities correlate significantly with identification of recent purging. The statistics produced by SAS® are given and explained below:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4220.490	3821.722
SC	4227.888	4295.182
-2 Log L	4218.490	3693.722

R-Square 0.0426	Max-rescaled R-Square 0.1443

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	524.7684	63	<.0001
Score	890.3928	63	<.0001
Wald	3728397759	63	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DrugsCocaDay2	5	12.2072	0.0321
DrugsInhaDay2	5	18.6123	0.0023
DrugsMethLife1	5	25.2941	0.0001
SexForce1	1	61.3827	<.0001
TobacDays4	6	34.7640	<.0001
VehicleOtherD3	4	13.4986	0.0091
VehicleSelfD3	4	8.5285	0.0740
ViolMultWeap1	4	19.3492	0.0007
ViolUnsafe2	4	11.4630	0.0218
ViolThreatSP2	7	64.7639	<.0001
ViolDamageSP2	7	19.3353	0.0072
ViolFightSP3	7	25.2556	0.0007
ViolInjury3	4	16.9803	0.0020

The maximum re-scaled R-square value for this model suggests that the risky behaviors identified above contribute to only 14.4% of the variability in identification of recent purging. Analysis of Maximum Likelihood Estimates and Odds ratio scores indicate that students who indicated less participation in DrugsCocaDay2, DrugsInhaDay2, DrugsMethLife1, VehicleOtherD3, VehicleSelfD3, ViolThreatSP2, ViolDamageSP2, ViolInjury3and ViolFightSP3 were significantly more likely to deny recent purging than students who identified greater participation in these variables. More specifically, students who denied participation in these activities had the greatest likelihood of denying having ever used purging to lose weight while students who indicated greater participation were more likely to identify these attempts regardless of number of times of usage/participation. On the other hand, students who identified less participation in ViolDamageSP2 and ViolFightSP3 were significantly more likely to identify recent purging than peers who identified greater participation in these activities. Similar results to the former were found for binary variables. SexForce1 indicated that students who had been forced to have sexual intercourse when they did not want to were also 3 times more likely to identify having used purging to lose weight recently. As with depression and suicidal ideation, this model suggests that when looking at purging to lose weight as a dependent variable, researchers and clinicians must also consider unsafe risky behaviors and environments that could put these students at risk for this mental health issue.

Latent Class Analysis

Through this paper, we can argue that using a factor analysis is effective in exploring a deeper relationship between different variables within the same categories of a survey analysis. However, another question arises as to just how these variables interact with each other and if there are latent variables that could be derived from these interactions. How is the specific organization of responses causing the different variables to load high in certain factors and how do these responses interact with each other? How are the participant responses to different variables causing the factors to be different? What is causing them to be similar? Is it possible that the interactions between our observed variables are actually creating unobserved (or latent) variables? What makes up these variables and how can we utilize them? One way we can explore these questions is through a latent class analysis. A latent class analysis is a concept used widely in the clinical sciences as it enables researchers to explore the relationship between observed (measured and/or discrete) variables and suggested latent variables that can be derived by the interactions of existing observed variables. There is no procedure within SAS that explores this type of analysis. However, a paper written by David S. Thompson and presented at SUGI 31 explored using PROC CATMOD as an option to perform a latent class analysis. This paper was quite comprehensive and explored the rationale behind using CATMOD (mainly given that it has multivariate loglinear modeling capabilities) as the procedure through which a latent class analysis would be able to be executed. To execute a latent class analysis through this route, one's code ends up looking something like this:

```
ods output
anova=mlr MaxLikelihood=iters estimates=mu covb=covb;
proc catmod data=YRBS_Total_LCA order=data;
       weight weight;
       model a*b*c*d*x =  response  / wls covb addcell=.1;
       loglin a b c d a*x b*x c*x d*x;
run;

data YRBS_Total_LCA;

       /*read in estimated model parameters (lambda1-lambda9) that were output
       from PROC step,
       then restructured in intervening data step.*/
       set mu;

       /*vector of CATMOD's loglinear parameter estimates*/
       array mu [9] lambda1-lambda9;

       /*vector of variable names*/
       array vars [4] a b c d;

       /*vector of conditional and LC probabilities*/
       array p [10] pa_x1 pb_x1 pc_x1 pd_x1 pa_x2 pb_x2 pc_x2 pd_x2 px1 px2;

       /*vector of joint probabilities*/
       array pjoint [2] piabcdx1 piabcdx2;
              do a=0 to 1;
                     do b=0 to 1;
                            do c=0 to 1;
                                   do d=0 to 1;
                                          do x=1 to 2;
                                                 do var=1 to 4;
                                                        value=vars[var];

       /*conditional probabilities for latent classes 1 and 2*/
              p[4*(x-1)+var]
              = exp(mu[var]*(-1)**value + mu[var+5]*(-1)**(value+x))

              /(exp(mu[var]*(-1)**value + mu[var+5]*(-1)**(value+x))
              +exp(-mu[var]*(-1)**value - mu[var+5]*(-1)**(value+x)));
```

```
/*latent class probabilities*/
      p[8+x]=exp(mu[5]*(-1)**(x-1))
      / (exp(mu[5]*(-1)**(x-1)) + exp(-mu[5]*(-1)**(x-1)));

/*joint probabilities for each class*/
      pjoint[x]=p[4*x-3]*p[4*x-2]*p[4*x-1]*p[4*x]*p[8+x];

/*unconditional predicted response probabilities (across both classes)*/
      piabcd=piabcdx1+piabcdx2;

/*Posterior probabilities (pix1_abcd) are the probabilities that an
individual
resides in latent class X=t, given observed responses A,B,C, and D*/
      pix1_abcd=piabcdx1/piabcd;
      pix2_abcd=piabcdx2/piabcd;
            if x=2 and var=4 then output;
                  end;
                        end;
                              end;
                                    end;
                                          end;
                                                end;
      drop lambda1-lambda9 value var x;
run;
```

Given the complexity of this option, this can prove quite daunting to execute; therefore, an alternative procedure designed specifically for an LCA could prove most beneficial. In response to the continuing need for an LCA-specific procedure, the Methodology Center at PennState University set to work on creating a more user friendly procedure to execute both a latent class analysis as well a latent transition analysis. The PennState LCA – LTA program can be included as an add-on to Base SAS® 9.x and associated SAS® Enterprise Guide. This program is available at http://methodology.psu.edu with instructions as to how to install and run it. The code used ends up looking something like this:

```
proc lca data=YRBS_Total_LCA;
      nclass 2;
      items MoodDep1 MoodConsiderS2 MoodPlanS3;
      categories 2 2 2;
      seed 861551;
run;
```

Given that this paper is written primarily for a factor analysis, a latent class analysis of the data presented in this paper would be unnecessary; however, this could prove a beneficial addition to any factor analysis in which the relationship between the factors and any possible latent variables would warrant exploration.

Conclusion

In conclusion, utilizing a factor analysis to gain a more thorough understanding as to the relationship and structure of survey variables can be a useful preliminary step to any logistic regression analysis. Considering the negative impact of multicollinearity, identifying variables that load into the same factors can prove beneficial in weeding out unnecessary variables that could negatively affect the validity and reliability of a logistic regression model. Further exploration of the variables and factors through a latent class analysis could prove even more beneficial as latent variables could be discovered and the potential of a more thorough understanding of the relationship between the different contributing variables/factors and their impact on the dependent/observed variable could be achieved.

This study was intended as an exercise on the use of factor analysis with logistic regression in a large survey sample of social science data, however, the significant results of the study warrant some discussion. In conclusion, this study supports the idea that youth participation in risky behaviors is correlated significantly with various possible mental illness indicators. These indicators were also identified as strong representations of the proposed risky behaviors through factor analysis. This study also found that overall, students who were female and in high school during their late tweens and early teens were at a greater risk for displaying mental illness characteristics than other students. The results of this study should encourage further research into the contributing factors of mental illness and youth participation in risky behaviors. If we can further strengthen our understanding of the connection between these two entities, proper preventative programs and therapies can be developed to assist students with various mental and physical health problems as they arise.

The limitations of this study stem from the fact that it is a secondary analysis based on a nationally distributed sample. The researcher could not be sure how the data was randomized and therefore, weights recalculated by the Center for Disease Control and Prevention were used in the analysis.

Acknowledgments

This author would like to acknowledge the Center for Disease Control and Prevention for their efforts in gathering the Youth Risk Behavior Surveillance System data. This author would also like to thank the CDC for making the national data available for usage online for research and educational purposes.

References

Allison, Paul D. 2012. *Logistic Regression Using SAS® : Theory and Application, Second Edition,* Cary, NC: SAS® Institute Inc.

Allison, Paul D. (2014, March). *Measures of Fit for Logistic Regression.* Paper presented at SAS® Global Forum 2014, Washington, D.C.

Center for Disease Control and Prevention (2014). Software for Analysis of YRBS Data. From http://www.cdc.gove/yrbs (accessed June, 2014).

Center for Disease Control and Prevention (2014). Combining YRBS data across years and sites. From http://www.cdc.gove/yrbs (accessed June, 2014).

Center for Disease Control and Prevention (2014). Methodology of the YRBSS. From http://www.cdc.gove/yrbs (accessed June, 2014).

Child, D. (1990). The essentials of factor analysis, second edition. London: Cassel Educational Limited.

Field, A., & Miles, J. (2012). *Discovering Statistics Using SAS®*, Thousand Oaks, CA: Sage Publications.

Introduction to SAS®. UCLA: Academic Technolgy Services, Statistical Consulting Group. From http://www.ats.ucla.edu/stat/sas/notes2/ (accessed August, 2012).

Korn, EL and Graubard, BI (1999). Analysis of Health Surveys. John Wiley & Sons, New York. p. 209-211.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A., & Collins, L. M. (2013). *PROC LCA & PROC LTA users' guide* (Version 1.3.0). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

O'Rourke, N., & Hatcher, L. 2013. *A Step-by-Step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling, Second Edition*, Cary, NC: SAS® Institute Inc.

PROC LCA & PROC LTA (Version 1.3.0) [Software]. (2013). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

Sanchez, M. E. "Analyzing the effects of adolescent risky behaviors on suicidal ideation." (2006). *Mathematics Theses*. Paper 21. http://digitalarchive.gsu.edu/math_theses/21

SAS® Institute Inc. 2008. SAS/STAT® 9.2 *User's Guide*. Cary, NC: SAS® Institute Inc.

Schreiber-Gregory, D. N. (2014, June). *Comparing Concatenation and Individual Analyses When Exploring Survey Data.* Paper presented at Wisconsin SAS Users Group Conference 2014, Milwaukee, WI.

Thompson, D. M. (2006). *Performing Latent Class Analysis Using the CATMOD Procedure*. Paper presented at SUGI 31, San Fransisco, CA.

YRBSS (2011). Retrieved April, 2012, from http://www.cdc.gove/yrbs

Contact Information

Your comments, questions, and suggestions are valued and encouraged. Contact the author at:

Deanna Schreiber-Gregory
North Dakota State University
(701) 799-6905
E-mail: d.n.schreibergregory@gmail.com or deanna.n.schreiber@ndsu.edu