ILLINOIS

# Sports Participation in the US:

## A Statistical Analysis

Luis Steven Lin

STAT 448
Group 9
12-12-12

# Contents

# List of Tables

## List of Figures

# 1. INTRODUCTION

## 1.1 Background

The 2010 American Time Use Survey (ATUS) was conducted by United States Bureau of Labor Statistics. The survey collects demographic data and information on the amount of time people spent doing various activities in 2010. Respondents were interviewed only once and were randomly selected from a subset of households that have completed their final month of interviews for the Current Population Survey.

The dataset was obtained from the Inter-university Consortium for Political and Social Research (ICPSR) website. The ICPSR is an international consortium of about 700 academic institutions and research organizations that provides access to more than 500,000 files of research in the social science, making ICPSR the world's largest archive of digital social science data.

The raw data file (30901-0007-Data.txt) is called the “Activity Summary File” and contains demographic information about respondents and the total time (in minutes) they spent doing each activity that day. The file is in fixed-column format, containing 170 variables (columns) and 13,260 records (rows). The record length is 3,309 and there is one record per unique respondent. There are no missing values in the raw data file. The activity variables correspond to the total number of minutes that each respondent spent doing each 6-digit activity. The 6-digit coding can be found in the documentation provided.

## 1.2 Problem Statement

The objective of this report was to apply analysis techniques to the 2010 American Time Use Survey (ATUS) to answer research questions related to the participation in sports activities. The main goal of the analysis was to determine if the participation is sports can be predicted by explanatory variables. More specifically, does the participation rate in sports differ by demographic groups? What factors are the most significant in determining sports participation? Is there any relationship between time spent on sports and time spent on other activities? The analysis techniques discussed in this report were aimed at answering these research questions.

# 2. METHODS

## 2.1 Analysis

Before performing a rigorous analysis, an exploratory analysis was conducted by computing descriptive statistics to provide a quantitative and graphical summary of the data, to detect patterns in the data, and to help avoid incorrect assumptions for the statistical analysis. In addition, the exploratory analysis was used to generate potential hypothesis for statistical testing. For the categorical variables, the descriptive statistics included frequencies and proportions for the various categories of the variables. For the continuous variables, the descriptive analysis consisted of computing central tendency (e.g. mean, dispersion (e.g. range and standard deviation) and distribution measures (e.g. skewness).

The next step was to perform an associative analysis to determine the presence of a relationship between variables. Spearman rank correlation was chosen to test the correlation because the method is non-parametric (e.g. no normality assumption) and does not assume linearity. The second method used to evaluate the association between variables was the Chi-square test. The chi-square tests measures if a significant relationship between the row and column variable exists, but they do not indicate the direction or strength of the relationship. More specifically, the Chi-square tests differences between observed and expected frequencies (Pearson and likelihood ratio are for general testing of association, while Mantel-Haenszel is for testing of trend).

For the people that participated in sports, an inferential analysis was conducted to detect significant differences in the time spent on sports among different groups of the categorical variables. Based on the descriptive analysis discussed previously, the assumption of normality is not reasonable. Thus, a non-parametric method for testing the differences in medians among groups was used. The chosen method was the Kruskal-Wallis test, which does not assume normality and is based on ranks of the data. The test assumes identically shaped and scaled distribution for each group, which seems to be a reasonable assumption based on the descriptive analysis.

The last step of the statistical analysis involved conducting a predictive analysis. Due to the dichotomous nature of the sports participation variable, a logistic regression model of this variable as a function of the explanatory variables was fitted to model the probability of observing the outcome of sports participation. Furthermore, a logistic model was selected because the previous analysis indicated that assumptions of normality, linearity and equal variance within each group might not be reasonable. The logistic model does not make these assumptions and there are no assumptions regarding the distribution of the independent variables.

The purpose of the predictive analysis was to discover relationships between the response variable and explanatory variables, and ultimately determine an adequate model capable of predicting the likelihood of participating in sports. The analysis started with the construction of an initial model by fitting the binary sports participation variable as a function of the predictor variables, which included the time activities as continuous variables and all demographic variables as categorical variables. The best model was found by stepwise selection, and a final model was chosen. Diagnostic for this model were performed and results were discussed.

The statistical analysis is summarized below in Figure 1 along with the SAS procedures utilized in each step of the analysis.

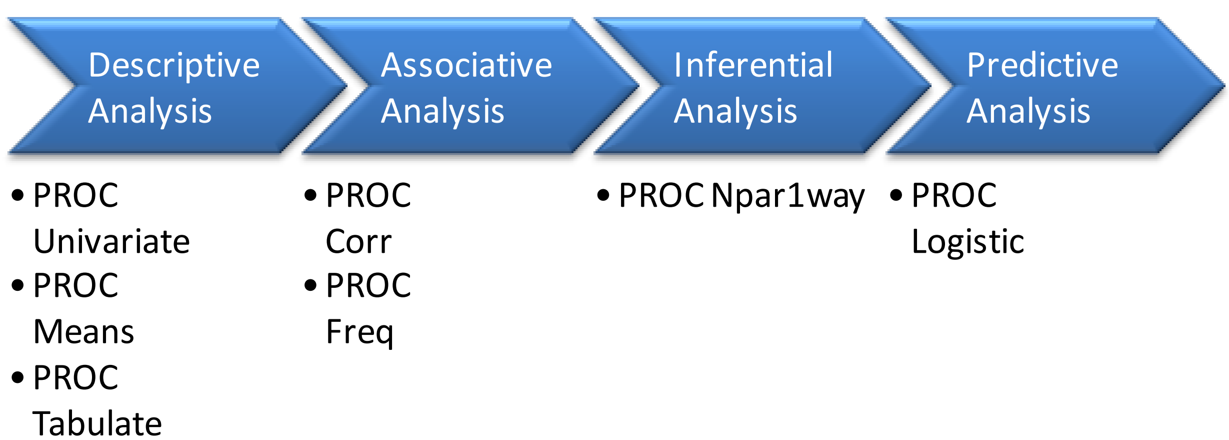



**Figure 1. Flow chart of the statistical analysis**

## 2.2 Variables

The time activities variables chosen are shown in Table 1. In order to answer the research question and fit a logistic regression model, the time spent on sports was converted to a binary variable which takes the value of 1 when the sports participation time is greater than zero and, takes the value of 0 when it is not. Thus, the binary variable represents the participation in sports. Additional variables were created by converting the time spent sleeping, eating and watching TV to categorical variable for the purpose of detecting group differences and comparing different models. The time spent smoking was also converted to a categorical variable which takes the value of 1 when the smoking time is greater than zero, and takes the value of 0 when it is not. Therefore, this binary variable indicates whether the participant smoked or did not smoke. The list of the categorical variables is shown in Table 2 in the following page.

**Table 1. Continuous Variables**

| Name | Description | Type | Values |
|---|---|---|---|
| **TSleep** | Time Sleeping | Continuous | 0-1395 min |
| **TEat** | Time Eating and Drinking | Continuous | 0-805 min |
| **TSocial** | Time Socializing & Communicating | Continuous | 0-940 min |
| **TSmoke** | Time Consuming Tobacco & Drug Use | Continuous | 0-300 min |
| **TTV** | Time Watching TV | Continuous | 0-1215 min |
| **TSPORTS** | Time Participating in Sports | Continuous | 0-777 min |

Table 2. Categorical Variables

| Name | Description | Type | Values |
|---|---|---|---|
| **Sex** | Sex of respondent | Nominal | Male<br>Female |
| **Age** | Age of respondent | Ordinal | 15 to 18 years<br>20 to 24 years<br>25 to 34 years<br>35 to 59 years<br>60 years and over |
| **Race** | Race of respondent | Nominal | White only<br>Black only<br>Native only<br>Asian only<br>Mixed or Other |
| **Hispanic** | Hispanic ethnicity of respondent | Nominal | Hispanic<br>Non-Hispanic |
| **Child** | Presence of child < 18 years in household | Ordinal | No Child<br>Child |
| **Status** | Labor for status of respondent | Nominal | Employed<br>Unemployed<br>Not in labor force |
| **Earnings** | Weekly earnings of respondent | Ordinal | No Income<br>Low Income<br>Medium Income<br>High Income |
| **Degree** | Education attainment of respondent | Ordinal | No High School Diploma<br>High School Diploma<br>Bachelors degree<br>Advanced degree |
| **Metropolitan** | Metropolitan location of respondent | Nominal | Metropolitan<br>Non-Metropolitan |
| **Smoke** | Tobacco or drug use | Nominal | No Smoke<br>Smoke |
| **Sleep** | Sleeping | Ordinal | Low Sleep<br>Medium Sleep<br>High Sleep |
| **Eat** | Eating and Drinking | Ordinal | Low Fat<br>Medium Fat<br>High Fat |
| **TV** | Watching TV | Ordinal | Low TV<br>Medium TV<br>High TV |
| **Sports** | Participation in sports | Nominal | No<br>Yes |

# 3. RESULTS

## 3.1 Exploratory Analysis

### *3.1.1 Time Activities*

The descriptive statistics for the continuous variables of the time activities are shown in Table 11. It can be seen that if non-participants of sports activities are considered, the mean time spent on sports is about 16 minutes and the median is 0 min. The table also shows that the mean is greater than the median for all time activities, suggesting positive-skewed distributions. In addition, it can be seen that the range is wide, with the maximum value being very large. For instance, for the time spent sleeping or watching television, the maximum is over 20 hours. Thus, this might suggest potential outliers in the data.

**Table 3. Descriptive statistics for continuous variables**

| *Variable* | *Label* | *N* | *Mean* | *Median* | *Minimum* | *Maximum* | *Std Dev* |
|---|---|---|---|---|---|---|---|
| TSPORTS | Time Participating in Sports | 6492 | 15.95 | 0.00 | 0.00 | 777.00 | 47.17 |
| TSLEEP | Time Sleeping | 6492 | 502.03 | 495.00 | 0.00 | 1395.00 | 129.66 |
| TEAT | Time Eating & Drinking | 6492 | 64.78 | 55.00 | 0.00 | 805.00 | 53.56 |
| TSOCIAL | Time Socializing & Communicating | 6492 | 29.93 | 0.00 | 0.00 | 940.00 | 71.38 |
| TTV | Time Watching TV | 6492 | 151.33 | 110.00 | 0.00 | 1215.00 | 163.33 |
| TSMOKE | Time Tobacco & Drug Use | 6492 | 0.36 | 0.00 | 0.00 | 300.00 | 5.88 |

### *3.1.2 Time Spent on Sports*

Figure 2 shows the overall participation in sports during weekdays, indicating that only 1188 out of the 6492 respondents participated in a sports activity. Focusing only on the people that participated in sports, the descriptive statistics of the time spent in sports were computed and are shown in Table 4 . The results indicate that participants in sports activity spent on average 87 minutes. The median time was much lower (60 minutes), suggesting that there are more values below the mean (i.e. more count in the tail than expected in a normal distribution). This is confirmed by the positive skewness (measure of asymmetry), which indicates that the mass of the time distribution is concentrated on the left (i.e. lower values of time).

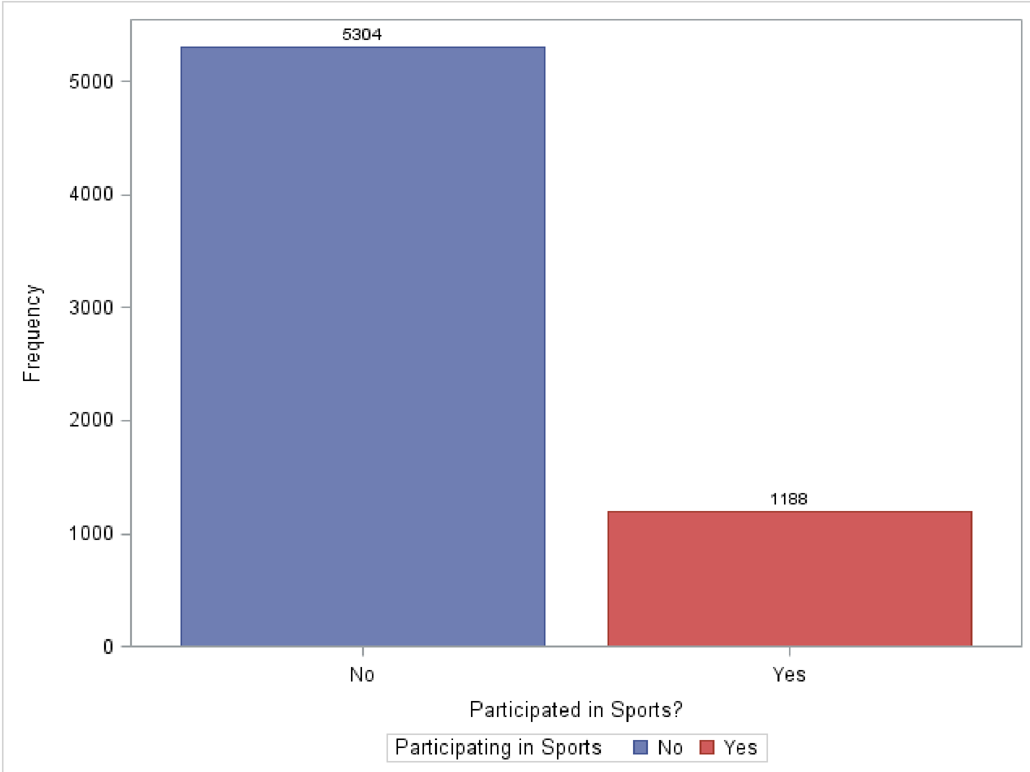



**Figure 2. Sports participation**

**Table 4. Descriptive statistics of times for sports participants**

| *Moments* | | | |
|---|---|---|---|
| *N* | 1188 | *Sum Weights* | 1188 |
| *Mean* | 87.1380471 | *Sum Observations* | 103520 |
| *Std Deviation* | 77.1970278 | *Variance* | 5959.3811 |
| *Skewness* | 2.67704398 | *Kurtosis* | 10.9659878 |
| *Uncorrected SS* | 16094316 | *Corrected SS* | 7073785.36 |
| *Coeff Variation* | 88.5916431 | *Std Error Mean* | 2.23971294 |

| *Basic Statistical Measures* | | | |
|---|---|---|---|
| *Location* | | *Variability* | |
| *Mean* | 87.13805 | *Std Deviation* | 77.19703 |
| *Median* | 60.00000 | *Variance* | 5959 |
| *Mode* | 60.00000 | *Range* | 772.00000 |
| | | *Interquartile Range* | 69.50000 |

The null hypothesis for the tests of normality (e.g. sample came from a normal distribution) is rejected at the 0.05 significance level (Table 5), suggesting that the distribution is of the time spent on sports is not normal. These results are graphically represented in Figure 3 which shows the histogram and probability plot.

**Table 5. Test for normality**

| *Tests for Normality* | | | | |
|---|---|---|---|---|
| *Test* | *Statistic* | | *p Value* | |
| *Shapiro-Wilk* | *W* | 0.748136 | *Pr < W* | <0.0001 |
| *Kolmogorov-Smirnov* | *D* | 0.203227 | *Pr > D* | <0.0100 |
| *Cramer-von Mises* | *W-Sq* | 13.75046 | *Pr > W-Sq* | <0.0050 |
| *Anderson-Darling* | *A-Sq* | 77.54448 | *Pr > A-Sq* | <0.0050 |

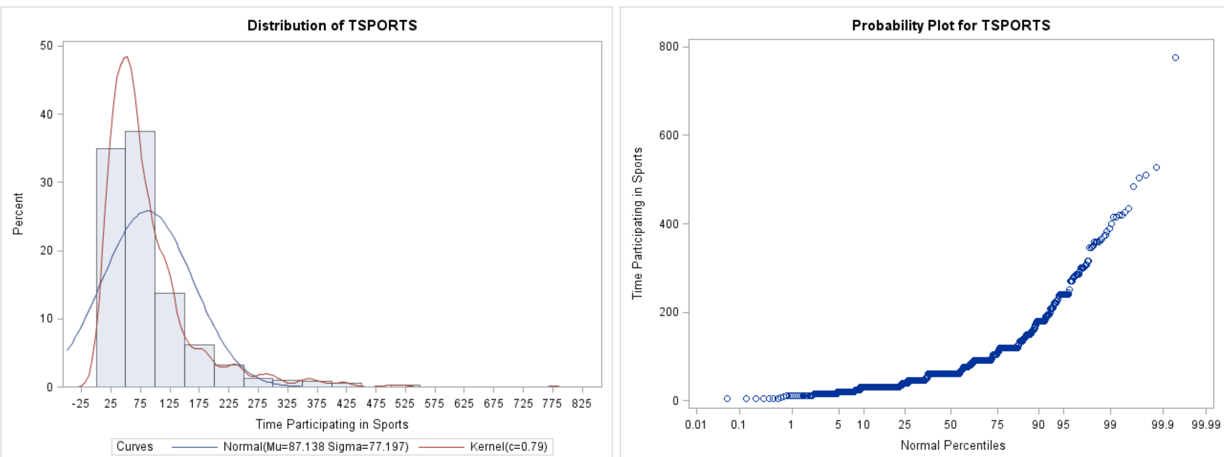



**Figure 3. Histogram and probability plot of sports time of participants**

### *3.1.3 Time Spent on Sports by categorical variables*

Histograms showing the distribution of the time for sports participants by the levels of each categorical variable were also constructed. Figure 4 shows the sports time distribution by sex of the participant. It can be seen that the distributions do not appear to be normal but the shape of the distributions of the sports time for men and women are very similar. This observation was consistent for the rest of the categorical variables (see Appendix A for details).

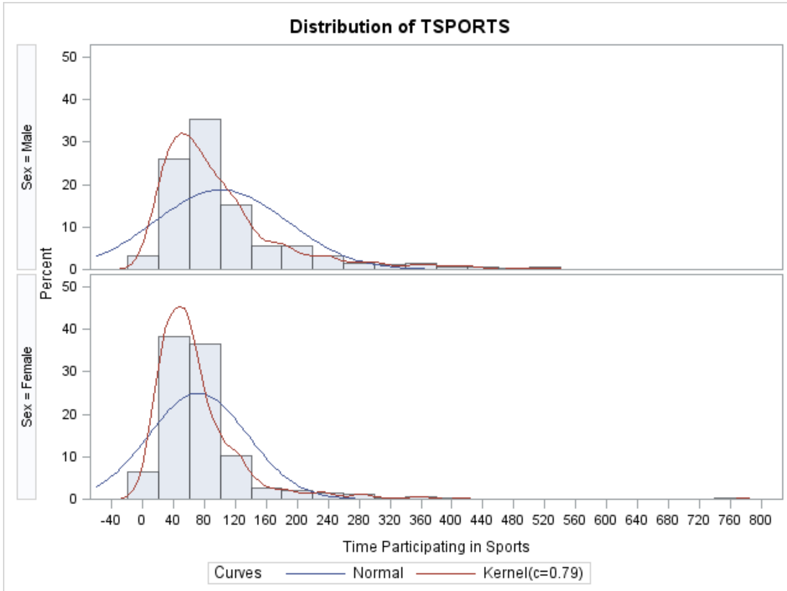



Figure 4. Histogram of TSPORTS by sex

For a further analysis, the time activities were broken down by categorical variables. For example, Figure 5 shows the mean time spent on activities by sex for all participants and non-participants of the activities. The plot suggests that there seems to be a significant difference between males and females regarding time spent watching TV and playing sports, with the latter one being almost twice of that of females for males.

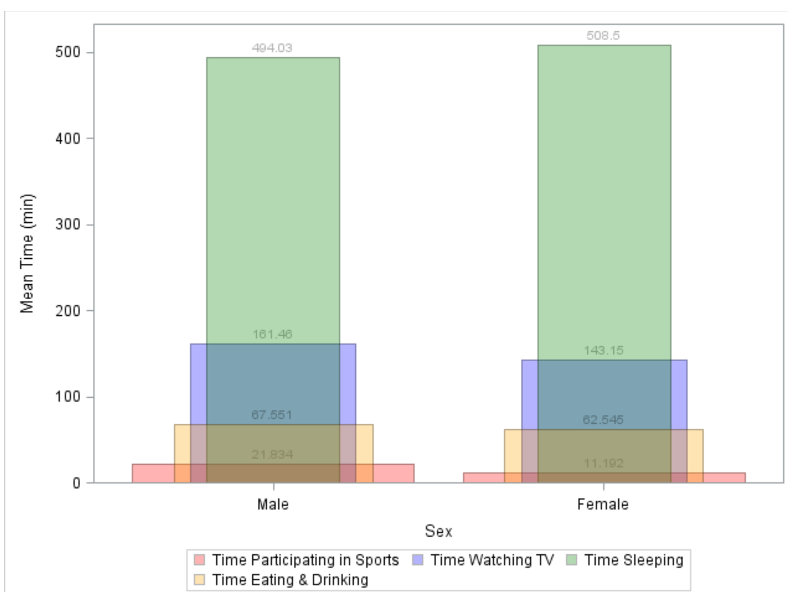



Figure 5. Time activities by sex

Focusing only on the people that participated in sports, the time spent on sports was further broken down by the rest of the categorical variables by constructing boxplot and cross-tabulations. The boxplot (Figure 6) and cross-tabulation (Table 6) of the time spent on sports by education and labor force status show that there seems to be a difference in the mean times and variation of the groups for both categorical variables.

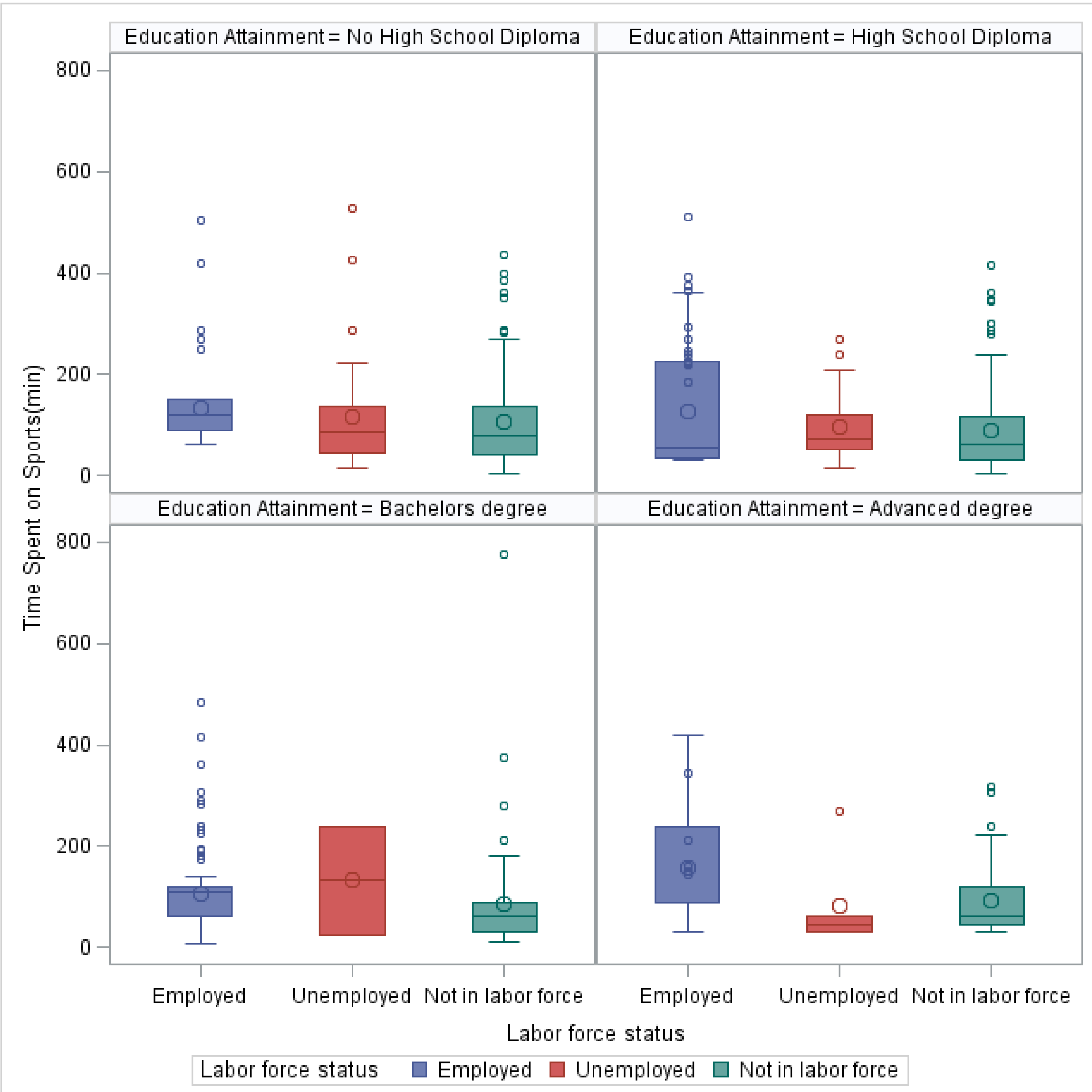



Figure 6. TSPORS by labor force and education

Table 6. TSPORTS by labor force and education

| | | Mean | Std | N |
|---|---|---|---|---|
| *Education Attainment* | *Labor force status* | | | |
| *No High School Diploma* | *Employed* | 111.55 | 92.21 | 56 |
| | *Unemployed* | 116.84 | 105.59 | 37 |
| | *Not in labor force* | 106.37 | 90.68 | 113 |
| *High School Diploma* | *Employed* | 82.12 | 70.77 | 287 |
| | *Unemployed* | 104.44 | 74.05 | 45 |
| | *Not in labor force* | 87.05 | 75.46 | 182 |
| *Bachelors degree* | *Employed* | 77.60 | 66.12 | 215 |
| | *Unemployed* | 95.91 | 64.20 | 11 |
| | *Not in labor force* | 85.19 | 109.15 | 63 |
| *Advanced degree* | *Employed* | 71.20 | 58.75 | 127 |
| | *Unemployed* | 80.33 | 93.67 | 6 |
| | *Not in labor force* | 91.00 | 67.16 | 46 |

Figure 7 and Table 7 indicate that the mean sports time seems to differ across different races, but not between metropolitan statuses. The boxplot also shows the high variability in the native only group due to the small sample of this group.

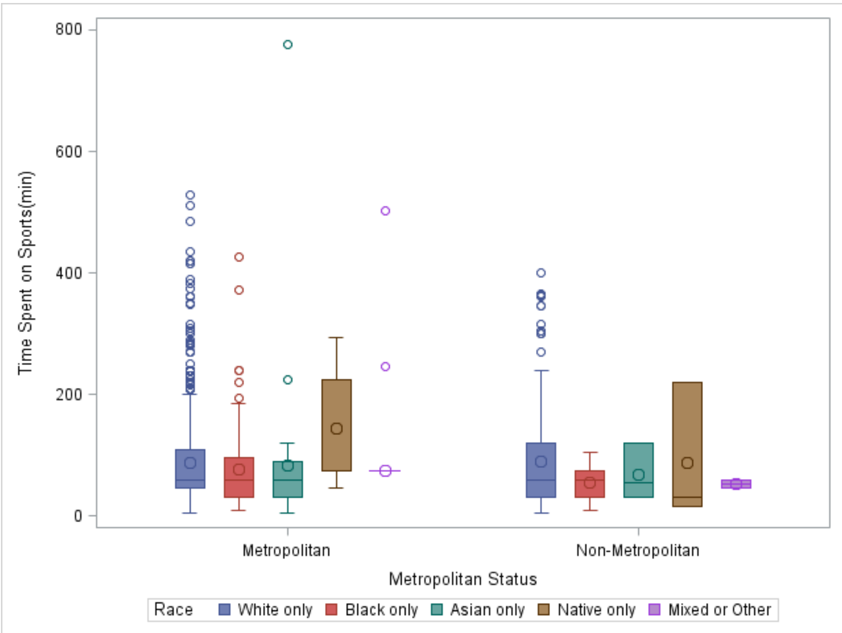



Figure 7. TSPORTS by metropolitan status and race

Table 7. TSPORTS by metropolitan status and race

| | | Mean | Std | N |
|---|---|---|---|---|
| *Metropolitan Status* | *Race* | | | |
| *Metropolitan* | *White only* | 88.31 | 74.58 | 800 |
| | *Black only* | 77.21 | 63.69 | 131 |
| | *Native only* | 142.80 | 109.97 | 5 |
| | *Asian only* | 82.66 | 115.31 | 44 |
| | *Mixed or Other* | 105.62 | 114.28 | 21 |
| *Non-Metropolitan* | *White only* | 90.24 | 82.56 | 165 |
| | *Black only* | 53.69 | 31.61 | 13 |
| | *Native only* | 88.33 | 114.27 | 3 |
| | *Asian only* | 68.33 | 46.46 | 3 |
| | *Mixed or Other* | 45.00 | 15.00 | 3 |

Looking at the sports time broken down by age and sex variables, Figure 8 and Table 8 suggest that seems to be a difference in the mean times of the groups for both categorical variables. For the sex variable, the mean and variation in time spent by males seems to be larger than that for females.

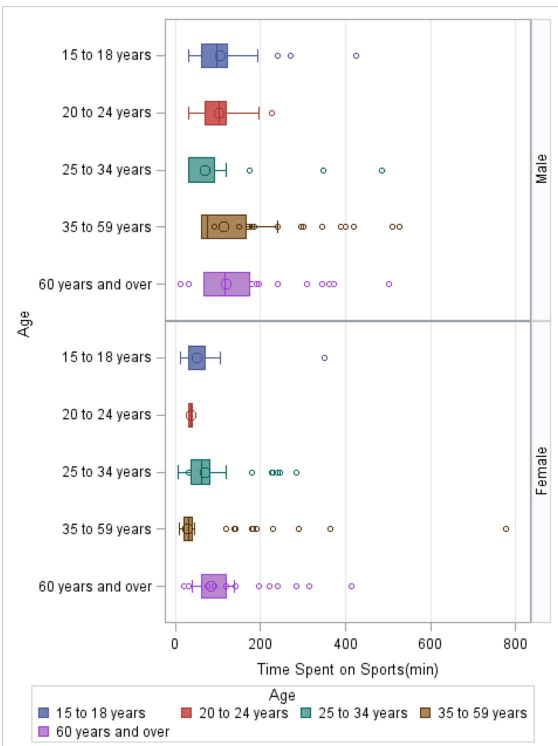



Figure 8. TSPORTS by age and sex

Table 8. TSPORTS by age and sex

| | | Mean | Std | N |
|---|---|---|---|---|
| *Sex* | *Age* | | | |
| *Male* | *15 to 18 years* | 135.71 | 89.70 | 85 |
| | *20 to 24 years* | 110.97 | 61.05 | 37 |
| | *25 to 34 years* | 95.51 | 74.15 | 87 |
| | *35 to 59 years* | 94.84 | 86.19 | 269 |
| | *60 years and over* | 94.27 | 87.44 | 147 |
| *Female* | *15 to 18 years* | 96.91 | 70.15 | 45 |
| | *20 to 24 years* | 79.64 | 63.31 | 14 |
| | *25 to 34 years* | 73.11 | 55.52 | 85 |
| | *35 to 59 years* | 66.04 | 66.11 | 253 |
| | *60 years and over* | 71.10 | 61.62 | 166 |

Figure 9 and Table 9 indicate that the mean sports time seems to differ slightly across different earnings levels, but not between Hispanic and non-Hispanic ethnicity. Regarding the smoke and child categorical variables (and), there seems to be a difference in the groups for both categorical variables, with the smoking group having a higher mean and child group having a lower mean. In all the boxplots shown above, a potential outlier with time spent on sports activities of about 800 min was detected.

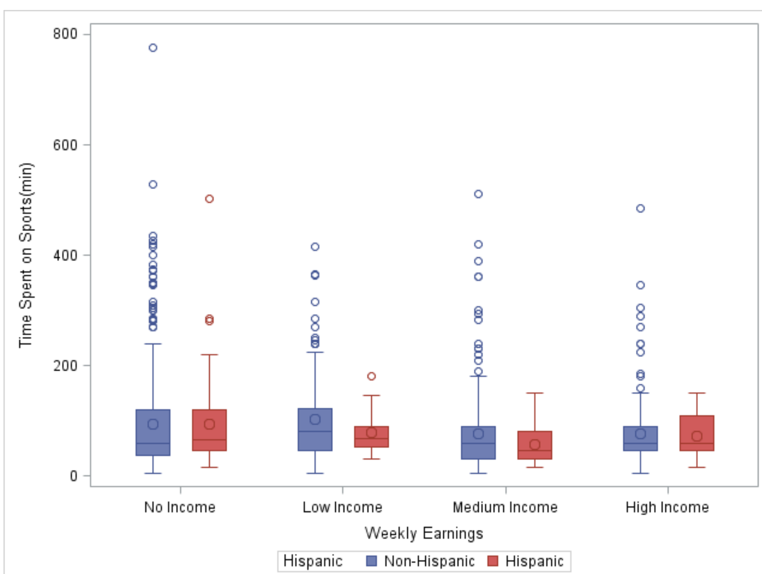



Figure 9. TSPORTS by earnings and race

Table 9. TSPORTS by earnings and race

| | | Mean | Std | N |
|---|---|---|---|---|
| *Weekly Earnings* | *Hispanic* | | | |
| *No Income* | *Hispanic* | 92.77 | 74.20 | 81 |
| | *Non-Hispanic* | 94.49 | 86.51 | 503 |
| *Low Income* | *Hispanic* | 78.25 | 40.76 | 20 |
| | *Non-Hispanic* | 102.08 | 81.35 | 116 |
| *Medium Income* | *Hispanic* | 57.08 | 35.14 | 36 |
| | *Non-Hispanic* | 76.94 | 70.78 | 245 |
| *High Income* | *Hispanic* | 72.27 | 40.33 | 11 |
| | *Non-Hispanic* | 75.99 | 61.54 | 176 |

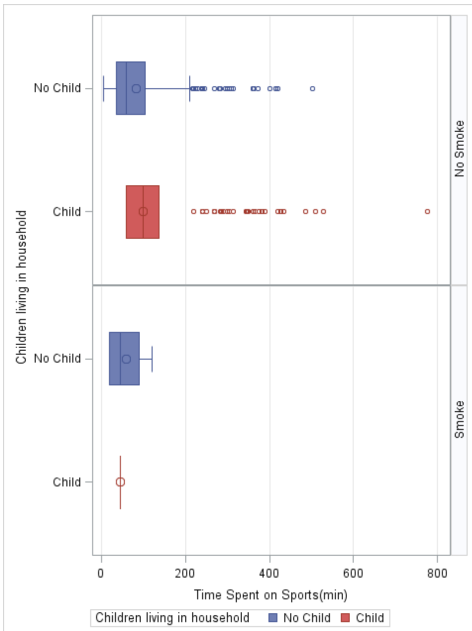



Figure 10. TSPORTS by smoke and child

Table 10. TSPORTS by smoke and child

| | | Mean | Std | N |
|---|---|---|---|---|
| *Tobacco & Drug Use* | *Children living in household* | | | |
| *No Smoke* | *No Child* | 82.59 | 70.30 | 635 |
| | *Child* | 93.02 | 84.71 | 543 |
| *Smoke* | *No Child* | 59.29 | 40.25 | 7 |
| | *Child* | 50.00 | 8.66 | 3 |

### *3.1.4 Participation in Sports by categorical variables*

The frequency distributions of the categorical variables are shown in Table 11 in the next page. Some features that can be pointed out are the low frequency of native only and smokers. The cross-tabulation of the participation rates in sports and the categorical variables is discussed in the next section. In this section, several plots of the number of participants in sports by categorical variables were constructed to continue the explorative analysis of the data. Note that the plots only show the number of participants rather than the proportion of participants, which is addressed in the following sections. Thus, no inferences on differences in sports participation between groups can be made in this section.

Figure 11 shows the participation in sports by age and sex. It can be seen that the relative number of male and females participating in sports is about the same except for the younger groups, in which there are more males than females participating based on the sample. The plot also shows that there are more observations for the older groups in the sample.

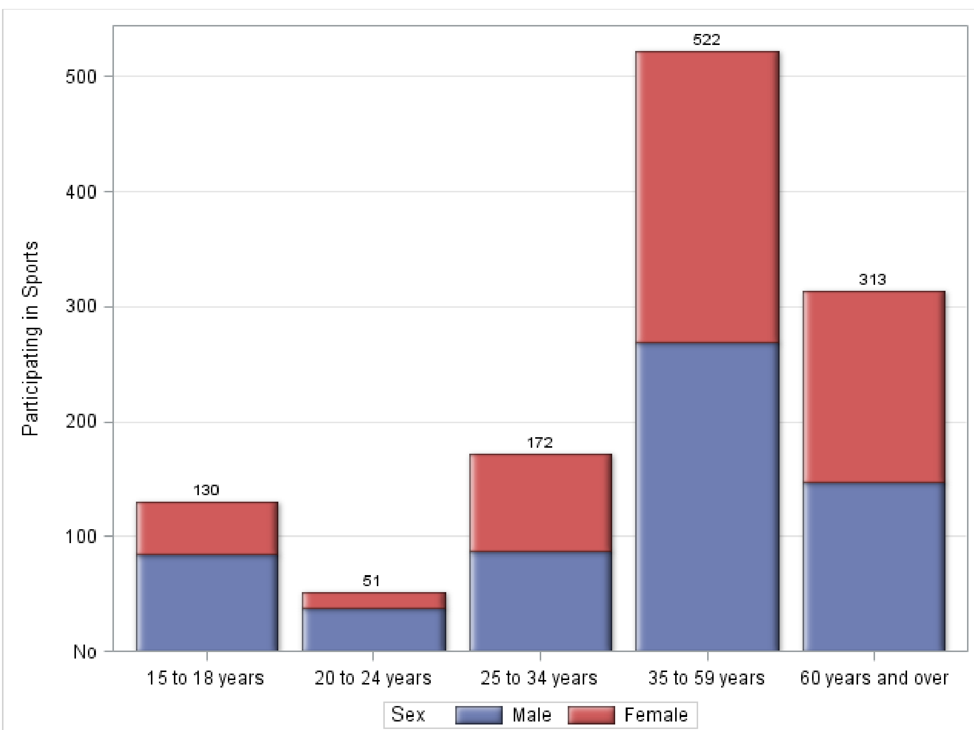



**Figure 11. Participation in sports by sex and age**

Table 11. Frequencies of categorical variables

| Variable | Value | Frequency | Percent |
|---|---|---|---|
| **Sex** | Male | 2900 | 44.67 |
| | Female | 3592 | 55.33 |
| **Age** | 15 to 18 years | 358 | 5.51 |
| | 20 to 24 years | 321 | 4.94 |
| | 25 to 34 years | 1085 | 16.71 |
| | 35 to 59 years | 3064 | 47.20 |
| | 60 years and over | 1664 | 25.63 |
| **Race** | White only | 5127 | 78.97 |
| | Black only | 959 | 14.77 |
| | Native only | 62 | 0.96 |
| | Asian only | 229 | 3.53 |
| | Mixed or Other | 115 | 1.77 |
| **Hispanic** | Hispanic | 912 | 14.05 |
| | Non-Hispanic | 5580 | 85.95 |
| **Child** | No Child | 3456 | 53.23 |
| | Child | 3036 | 46.77 |
| **Status** | Employed | 3942 | 60.72 |
| | Unemployed | 452 | 6.96 |
| | Not in labor force | 2098 | 32.32 |
| **Earnings** | No Income | 3038 | 46.80 |
| | Low Income | 964 | 14.85 |
| | Medium Income | 1683 | 25.92 |
| | High Income | 807 | 12.43 |
| **Degree** | No High School Diploma | 1059 | 16.31 |
| | High School Diploma | 3386 | 52.16 |
| | Bachelors degree | 1314 | 20.24 |
| | Advanced degree | 733 | 11.29 |
| **Metropolitan** | Metropolitan | 5359 | 82.55 |
| | Non-Metropolitan | 1133 | 17.45 |
| **Smoke** | No Smoke | 6385 | 98.35 |
| | Smoke | 107 | 1.65 |
| **Sleep** | Low Sleep | 1971 | 30.36 |
| | Medium Sleep | 3280 | 50.52 |
| | High Sleep | 1241 | 19.12 |
| **Eat** | Low Eat | 1319 | 20.32 |
| | Medium Eat | 3554 | 54.74 |
| | High Eat | 1619 | 24.94 |
| **TV** | Low TV | 1567 | 24.14 |
| | Medium TV | 3600 | 55.45 |
| | High TV | 1325 | 20.41 |

Looking at the sports participation broken down by education and child variables, Figure 12 shows that the highest number of participants in sports in the sample was for the education attainment of high school diploma. The plot also shows that there are more sports participants in the sample with no child for people with at least a high school degree. As expected, for no high school degree, there are more participants in sports if there is a child present in the household because the participant is most likely to be a child.

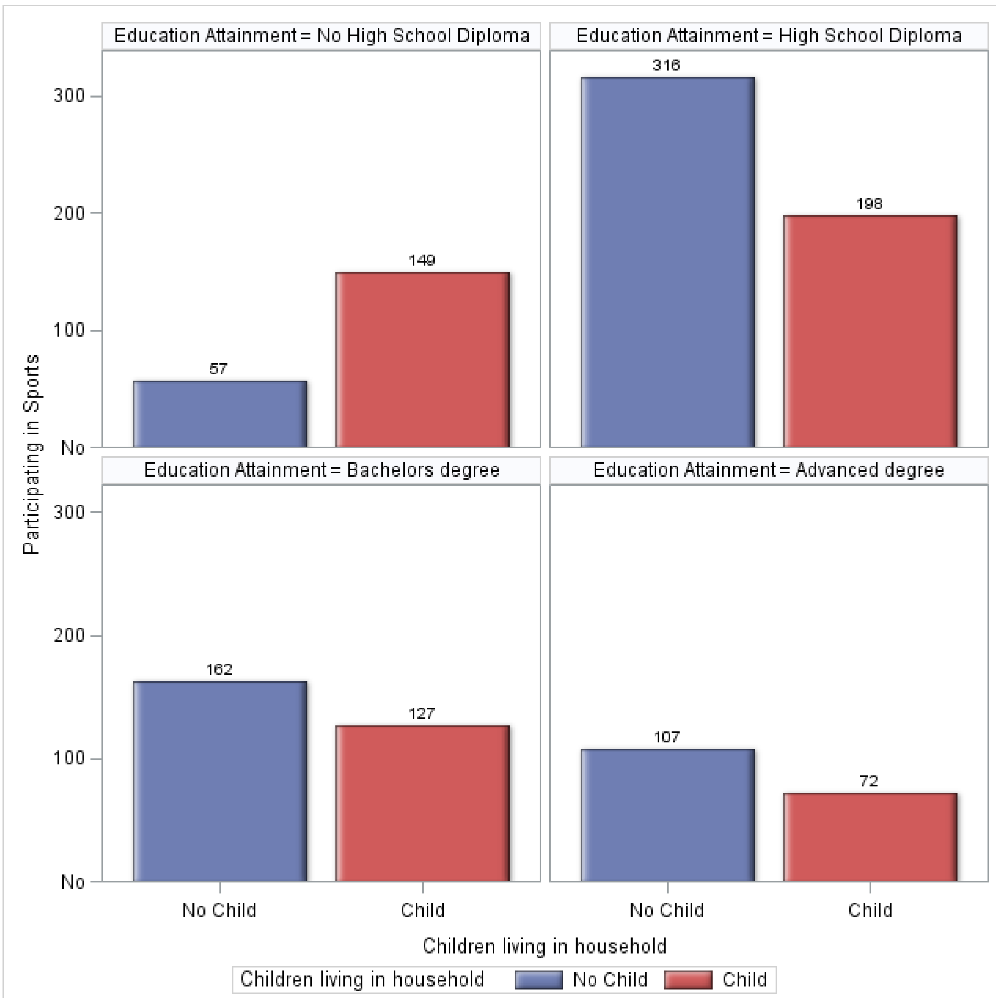



**Figure 12. Participation in sports by education and child**

Figure 13 shows that the sample contains more sports participants from metropolitan areas than non-metropolitan areas and this is consistent across different income levels. Figure 14 indicates than the sample has only a few sports participants who are smokers and that the sample contains more sports participants of the race white only.

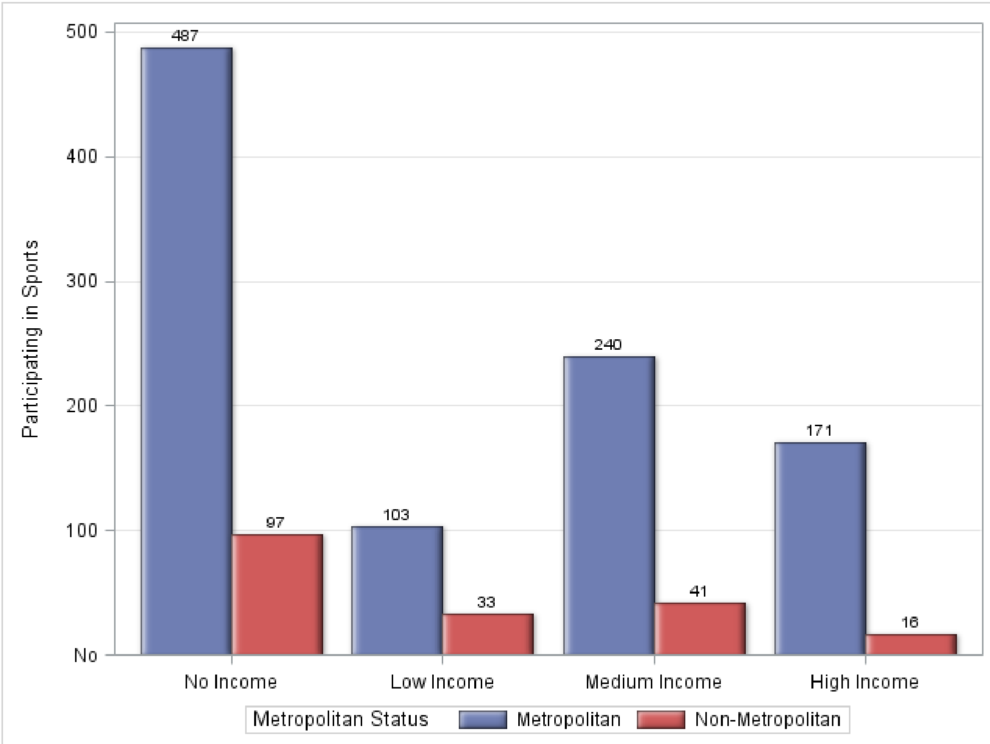



**Figure 13. Participation in sports by metropolitan status and earnings**

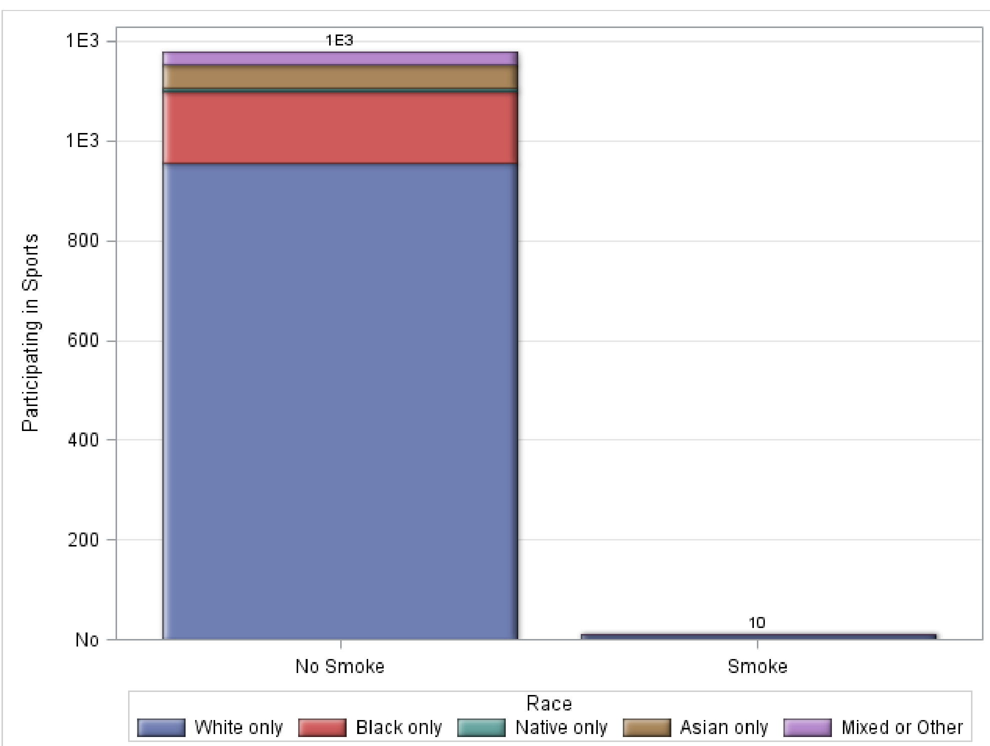



**Figure 14. Participation in sports by race and smoke**

Finally, Figure 15 shows that the relative sports participation distribution of the labor force status in the sample was about the same of Hispanic and non-Hispanic ethnicity. It can be seen that the sample contains more non-Hispanic observations and that there are more sports participants that are employed compared to not in the labor force or unemployed.

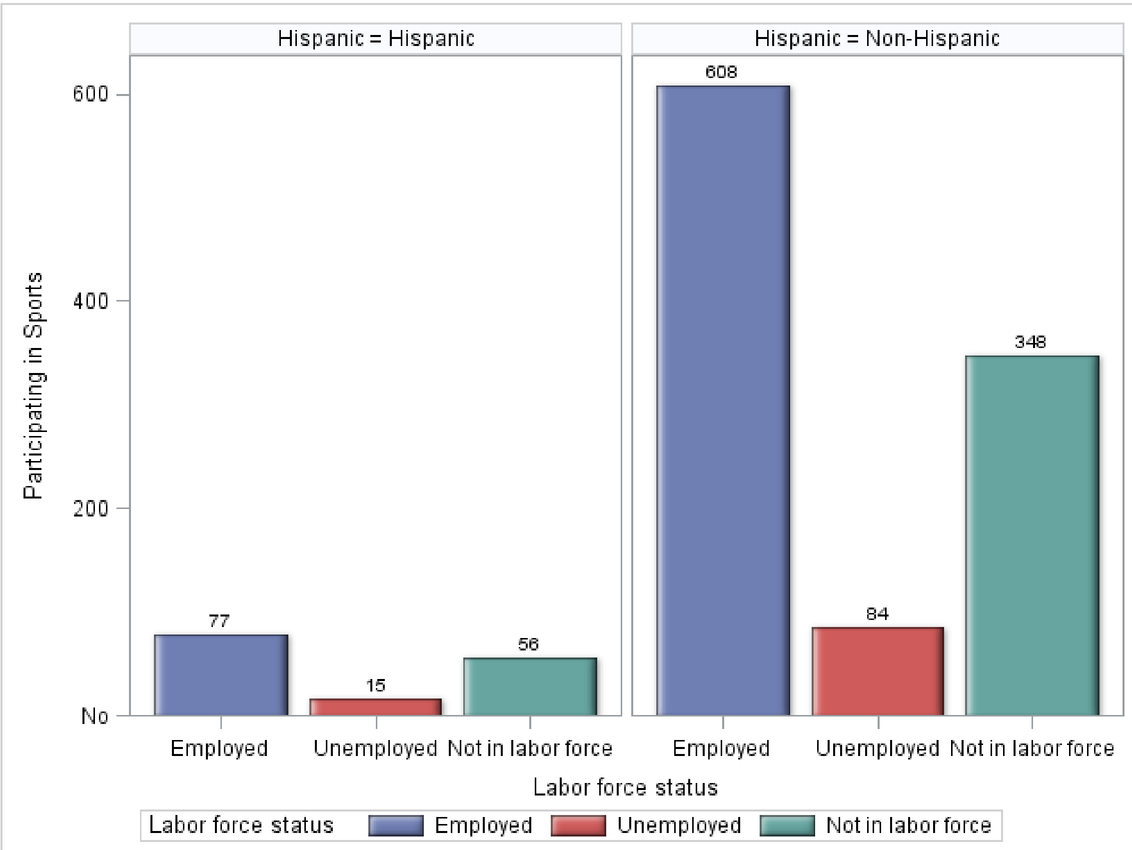



Figure 15. Participation in sports by labor force status and Hispanic ethnicity

## 3.2 Associative Analysis

The scatter plot (Figure 16) shown in the next page for the continuous variables indicates that there does not seem to be a linear relationship between variables and that the strength of the correlation is very weak or negligible. The Spearman correlation for the time activities is shown in Table 12 and suggests that the correlation between time spent on sports and time sleeping, eating and watching TV are statistically significant but the strength of the correlation is negligible, which might be due to the large sample.

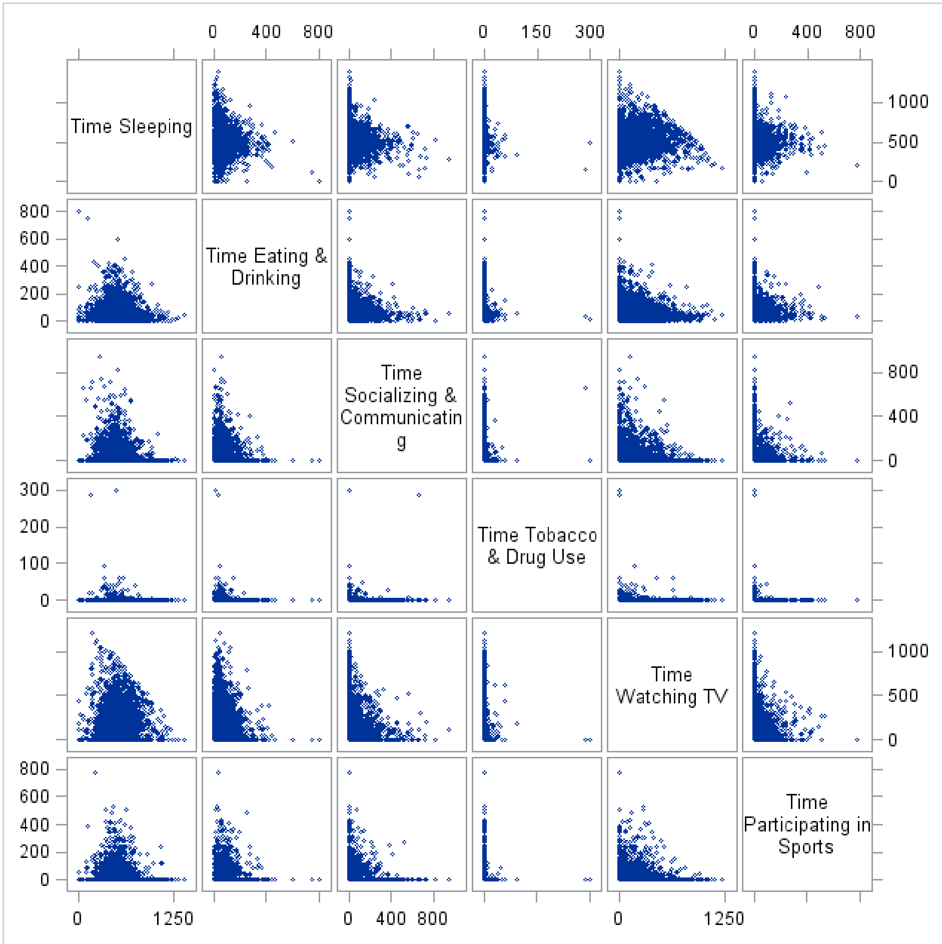



Figure 16. Scatter plot for continuous variables

Table 12. Correlation matrix for continuous variables

*Spearman Correlation Coefficients, N = 6492*
*Prob > |r| under H0: Rho=0*

| | *TSLEEP* | *TEAT* | *TSOCIAL* | *TTV* | *TSPORTS* |
|---|---|---|---|---|---|
| *TSLEEP*<br>Time Sleeping | 1.00000 | -0.04257<br>0.0006 | -0.05675<br><.0001 | 0.13185<br><.0001 | -0.03090<br>0.0128 |
| *TEAT*<br>Time Eating & Drinking | -0.04257<br>0.0006 | 1.00000 | -0.01473<br>0.2352 | -0.03596<br>0.0038 | 0.03679<br>0.0030 |
| *TSOCIAL*<br>Time Socializing & Communicating | -0.05675<br><.0001 | -0.01473<br>0.2352 | 1.00000 | -0.08391<br><.0001 | 0.02213<br>0.0746 |
| *TTV*<br>Time Watching TV | 0.13185<br><.0001 | -0.03596<br>0.0038 | -0.08391<br><.0001 | 1.00000 | -0.05331<br><.0001 |
| *TSPORTS*<br>Time Participating in Sports | -0.03090<br>0.0128 | 0.03679<br>0.0030 | 0.02213<br>0.0746 | -0.05331<br><.0001 | 1.00000 |

A similar result can be seen in the Spearman correlation for the dichotomous variables (Table 13), in which the correlation between participation in sports and the variables sex and smoke is statistically significant but the strength of the correlation is negligible. The sign of the correlation coefficient indicates the direction of the correlation. In this case, the negative signs of the coefficients between participation in sports and the variables sex and smoke make sense, since female and smokers tend to participate less in sports compared to males and non-smokers respectively.

**Table 13. Correlation matrix for continuous variables**

*Spearman Correlation Coefficients, N = 6492*
*Prob > |r| under H0: Rho=0*

| | *SEX* | *CHILD* | *HISPANIC* | *METRO* | *SMOKE* | *SPORTS* |
|---|---|---|---|---|---|---|
| *SEX*<br>Sex | 1.00000 | 0.03941<br>0.0015 | 0.01481<br>0.2329 | -0.01284<br>0.3009 | -0.02491<br>0.0448 | -0.07558<br><.0001 |
| *CHILD*<br>Children living in household | 0.03941<br>0.0015 | 1.00000 | -0.08264<br><.0001 | -0.01136<br>0.3602 | -0.00168<br>0.8925 | -0.00498<br>0.6880 |
| *HISPANIC*<br>Hispanic | 0.01481<br>0.2329 | -0.08264<br><.0001 | 1.00000 | 0.09080<br><.0001 | 0.02110<br>0.0891 | 0.02166<br>0.0810 |
| *METRO*<br>Metropolitan Status | -0.01284<br>0.3009 | -0.01136<br>0.3602 | 0.09080<br><.0001 | 1.00000 | 0.01948<br>0.1166 | -0.02082<br>0.0935 |
| *SMOKE*<br>Tobacco & Drug Use | -0.02491<br>0.0448 | -0.00168<br>0.8925 | 0.02110<br>0.0891 | 0.01948<br>0.1166 | 1.00000 | -0.02997<br>0.0157 |
| *SPORTS*<br>Participating in Sports | -0.07558<br><.0001 | -0.00498<br>0.6880 | 0.02166<br>0.0810 | -0.02082<br>0.0935 | -0.02997<br>0.0157 | 1.00000 |

A cross-tabulation for the participation in sports and the demographic variable sex is shown in Table 14. The cross-tabulation indicates that 21.55% of men and 15.67% of women participated in sports. In order to test if the difference in sports participation rates is statistically significant, chi-square tests were conducted and are shown in Table 15. The low p-values for the three tests indicate that the null hypothesis (i.e. there are approximately equal numbers of cases in each group) should be rejected in favor of the alternative hypothesis at the 0.05 level of significance. In other words, there is significant evidence of an association between participation in sports and sex of the participant (e.g. the sports participation rates are significantly different for men and women).

**Table 14. Cross-tabulation for sports and sex**

| Table of SPORTS by SEX | | | |
|---|---|---|---|
| *SPORTS(Participating in Sports)* | *SEX(Sex)* | | |
| *Frequency<br>Col Pct* | *Male* | *Female* | *Total* |
| *No* | 2275<br>78.45 | 3029<br>84.33 | 5304 |
| *Yes* | 625<br>21.55 | 563<br>15.67 | 1188 |
| *Total* | 2900 | 3592 | 6492 |

**Table 15. Chi-square test for sports and sex**

| *Statistic* | *DF* | *Value* | *Prob* |
|---|---|---|---|
| *Chi-Square* | 1 | 37.0811 | <.0001 |
| *Likelihood Ratio Chi-Square* | 1 | 36.8849 | <.0001 |
| *Continuity Adj. Chi-Square* | 1 | 36.6890 | <.0001 |
| *Mantel-Haenszel Chi-Square* | 1 | 37.0754 | <.0001 |
| *Phi Coefficient* | | -0.0756 | |
| *Contingency Coefficient* | | 0.0754 | |
| *Cramer's V* | | -0.0756 | |

The analysis was repeated for the rest of the demographic variables and the results are summarized in Table 16 (insignificant associations with sports participation are highlighted in red). The full details of the analysis can be found in Appendix B.

Table 16. Chi-Square Tests

| Variable | Chi-Square | Likelihood Ratio Chi-Square | Mantel-Haenszel Chi-Square |
|---|---|---|---|
| **SEX** | <.0001 | <.0001 | <.0001 |
| **AGE** | <.0001 | <.0001 | **0.1217** |
| **RACE** | **0.0353** | **0.0296** | **0.6637** |
| **HISPANIC** | **0.0810** | **0.0771** | **0.0810** |
| **CHILD** | **0.5381** | **0.5380** | **0.5381** |
| **STATUS** | 0.0241 | 0.0262 | 0.0494 |
| **EARNINGS** | <.0001 | <.0001 | 0.0050 |
| **DEGREE** | <.0001 | <.0001 | 0.0112 |
| **METRO** | **0.0855** | **0.0823** | **0.0856** |
| **SMOKE** | 0.0157 | 0.0090 | 0.0157 |
| **SLEEP** | 0.0006 | 0.0004 | 0.0717 |
| **TV** | 0.0002 | 0.0001 | <.0001 |
| **EAT** | 0.0221 | 0.0204 | 0.0179 |

## 3.3 Inferential Analysis

The results of Kruskal-Wallis test for determining significance difference in the time spent in sports (for the people that participated) between males and females is shown in Table 17 and Figure 17. The figure shows a boxplot for the scores by the sex variable, which seems to show a difference in the Wilcoxon score between men and women. The table indicates that the low p-value suggests that the null hypothesis (i.e. samples come from populations with identical locations) should be rejected at the 0.05 significance level. Thus, there is evidence that suggest at least one of the population medians differs from the others. In this case, this means that the median time spent on sports by men is significantly higher than that of the women.

Table 17. Kruskal-Wallis Test

| *Kruskal-Wallis Test* | |
|---|---|
| *Chi-Square* | 53.8647 |
| *DF* | 1 |
| *Pr > Chi-Square* | <.0001 |


Figure 17. Box plot of Wilcoxon scores for TSPORTS

A similar analysis was carried out for the rest of the demographic variables and the results are summarized in Table 18 (significant difference in sports time among groups is highlighted in green). The full details of the analysis can be found in Appendix C.

Table 18. Kruskal-Wallis Test Summary

| Variable | Pr > Chi-Square |
|---|---|
| **SEX** | **<.0001** |
| **AGE** | **<.0001** |
| **RACE** | 0.2530 |
| **HISPANIC** | 0.9058 |
| **CHILD** | **0.0313** |
| **STATUS** | **0.0036** |
| **EARNINGS** | **<.0001** |
| **DEGREE** | **0.0001** |
| **METRO** | 0.3248 |
| **SMOKE** | 0.1859 |
| **SLEEP** | 0.0630 |
| **TV** | **0.0014** |
| **EAT** | 0.1821 |

Note that the tests performed in the previous sections involve determining statistical significant differences (i.e. reject the null hypothesis), meaning that the differences are not likely due to chance or sampling variability. However, statistical significant differences can be found if the sample size is large enough and might not be practically significant.

## 3.4 Predictive Analysis

### *3.4.1 Model Selection*

An initial model was constructed by fitting the binary sports participation variable as a function of the predictor variables, which included the time activities as continuous variables and all demographic variables as categorical variables. The best model was found by stepwise selection, and a final model was chosen. The full output of the stepwise selection can be found in Appendix D. The stepwise selection summary with type 3 analysis of effects and ROC curves are shown below in Table 19 and Figure 18 respectively.

**Table 19. Stepwise Selection Summary**

*Summary of Stepwise Selection*

| *Step* | *Effect Entered* | *Effect Removed* | *DF* | *Number In* | *Score Chi-Square* | *Wald Chi-Square* | *Pr > ChiSq* | *Variable Label* |
|---|---|---|---|---|---|---|---|---|
| 1 | *AGE* | | 4 | 1 | 86.8509 | | <.0001 | Age |
| 2 | *DEGREE* | | 3 | 2 | 78.7494 | | <.0001 | Education Attainment |
| 3 | *SEX* | | 1 | 3 | 32.9711 | | <.0001 | Sex |
| 4 | *TTV* | | 1 | 4 | 18.9077 | | <.0001 | Time Watching TV |
| 5 | *STATUS* | | 2 | 5 | 13.6808 | | 0.0011 | Labor force status |
| 6 | *TSLEEP* | | 1 | 6 | 8.9586 | | 0.0028 | Time Sleeping |
| 7 | *CHILD* | | 1 | 7 | 4.5772 | | 0.0324 | Children living in household |

*Type 3 Analysis of Effects*

| *Effect* | *DF* | *Wald Chi-Square* | *Pr > ChiSq* |
|---|---|---|---|
| *SEX* | 1 | 38.1780 | <.0001 |
| *AGE* | 4 | 76.4926 | <.0001 |
| *CHILD* | 1 | 4.5722 | 0.0325 |
| *STATUS* | 2 | 18.0390 | 0.0001 |
| *DEGREE* | 3 | 61.3474 | <.0001 |
| *TTV* | 1 | 26.0760 | <.0001 |
| *TSLEEP* | 1 | 9.4614 | 0.0021 |

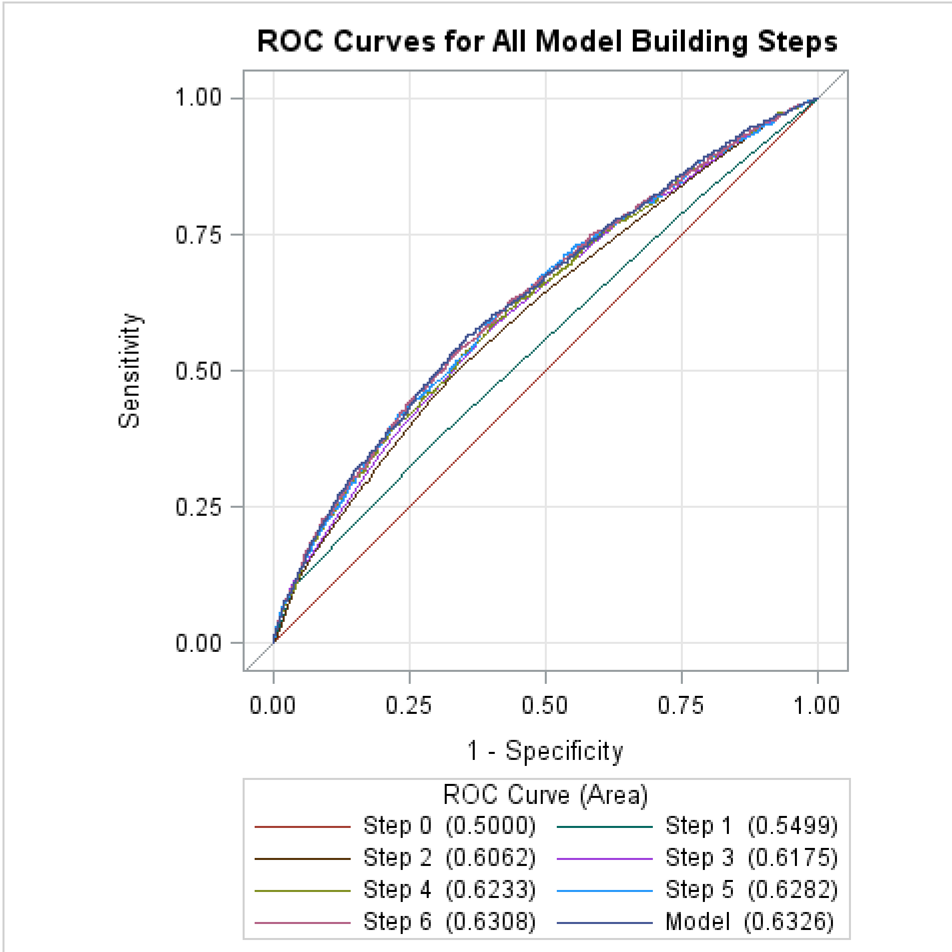



Figure 18. ROC curves for stepwise selection

The results show that based on the stepwise selection procedure, the best model is composed of 7 significant predictors of the 14 explanatory variables in the initial model. The selected demographic variables (categorical) are age, degree, sex, labor force status and presence of a child. The time activities variables (continuous) are sleeping and watching TV. The summary table and type 3 analysis (amount of variation that predictors add to the model given all other predictors are in the model) show consistent results, indicating that all selected predictors are significant at 0.05 level. The ROC curve shows that the area under the curve (AUC) is highest for the selected model. AUC is a measure of discrimination or predictive power, where 0.5 means that the model does not predict better than chance and 1.0 that the model discriminates perfectly. Thus, the selected model is predicting better than chance, but the AUC is not very high, suggesting that the model is barely acceptable and there is still room for improvement.

### 3.4.2 Model Fit

Measures of model fit include testing the global null hypothesis that the model coefficients are equal to zero, conducting a residual chi-square test, and performing a Hosmer-Lemeshow goodness of fit test. The results (Table 20) show that the global null hypothesis is rejected, indicating that at least one of the coefficients is different than zero. In addition, the residual chi-square test shows that the selected reduced model is adequate (null hypothesis that the reduced model is adequate is not rejected). The Hosmer-Lemeshow test shows that the model fits the data well (null hypothesis that there is no difference between observed and predicted values of the response variable is not rejected).

**Table 20. Goodness of fit tests**

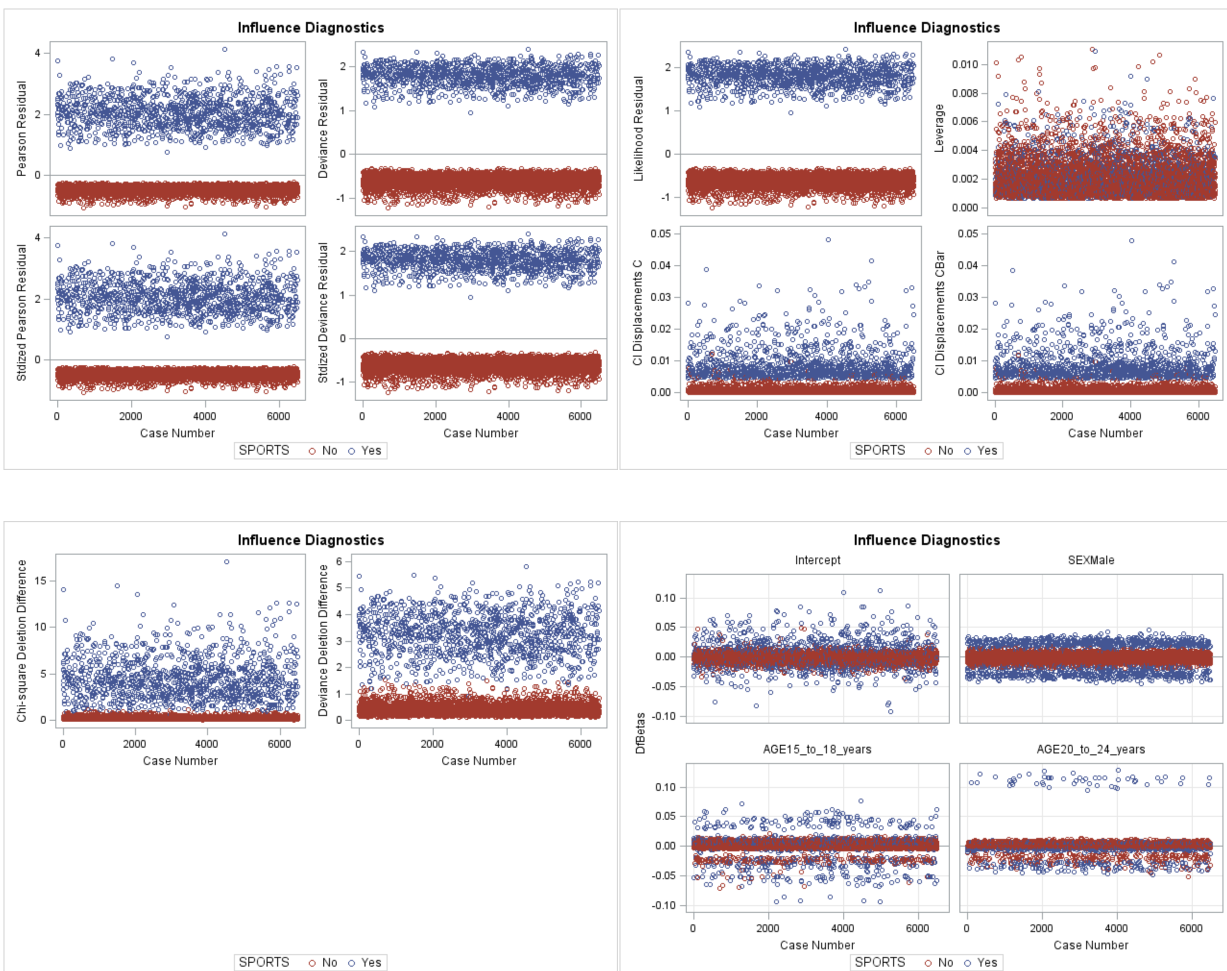
| *Testing Global Null Hypothesis: BETA=0* | | | |
|---|---|---|---|
| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
| *Likelihood Ratio* | 231.0671 | 13 | <.0001 |
| *Score* | 239.5551 | 13 | <.0001 |
| *Wald* | 225.3341 | 13 | <.0001 |

| *Residual Chi-Square Test* | | |
|---|---|---|
| *Chi-Square* | *DF* | *Pr > ChiSq* |
| 14.7921 | 12 | 0.2530 |

| *Hosmer and Lemeshow Goodness-of-Fit Test* | | |
|---|---|---|
| *Chi-Square* | *DF* | *Pr > ChiSq* |
| 8.6179 | 8 | 0.3755 |

### 3.4.3 Model Diagnostics

Diagnostics were conducted to assess the validity of the model. The assessment included influence diagnostics to identify highly influential points by considering differences between fitted and observed values. Predicted probability diagnostics and leverage diagnostics were also conducted. The diagnostic plots are shown below in Figure 19 . The residual in the influential plots fall along a horizontal band and also fall within the expected +/- 3 for the standardized residuals, indicating that there does not seem to be potential outlier or highly influential points. The predicted probability plots did not show any suspicious point not following the trends. The displacement measures and deviance deletion differences, which capture the effect of influential points in the coefficients, were not found to be a problem. Therefore, based on the diagnostics and goodness of fit, it was determined that the model was adequate.



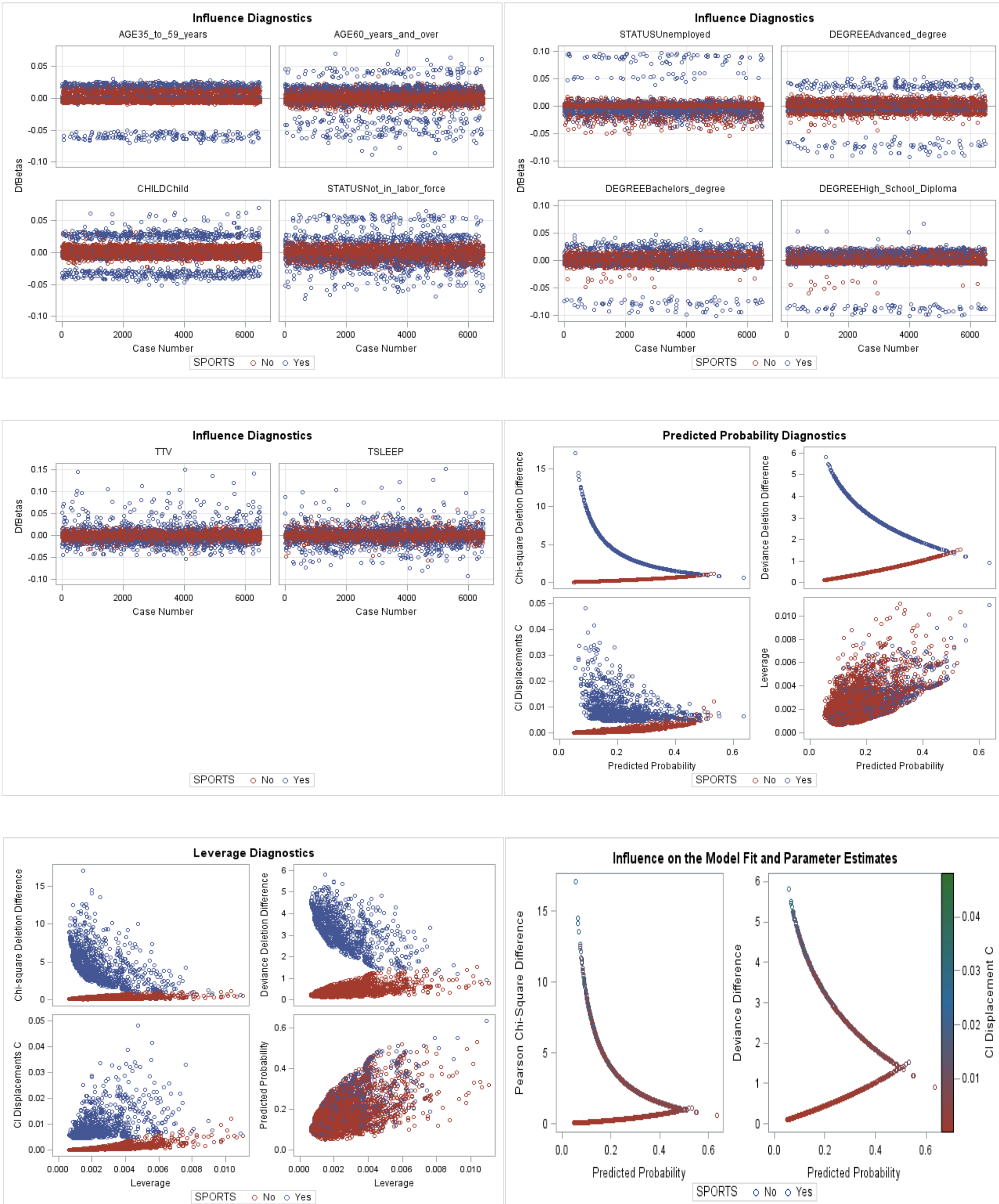



**Figure 19. Diagnostics Plots**

### *3.4.4 Model Discussion*

The coefficient estimates of the final model are shown below in Table 21. The maximum likelihood estimates are the coefficients estimates of the linear combination of the predictor variables that model the log odds of the response. For example, the coefficient of the parameter sex is 0.4151, which means that for one unit change in the predictor (e.g. sex), the difference in log-odds for the outcome is expected to change by 0.4151, given the other variables in the model remain constant. In this case, female was set as the reference level, so the difference in log-odds is expected to be 0.4151 units higher for males compared to females, given the other variables are held constant. The sign indicates whether the change in the predictor will increase (positive sign) or decrease (negative sign) the log odds. In this case, only the coefficients of child and time spent watching TV and sleeping are negative. The small coefficient of the latter two variables suggests a very small negative effect of these predictors on the log odds. The variables highlighted below in red indicate statistical insignificance of the corresponding coefficient terms.

**Table 21. Logistic model coefficient estimates**

*Analysis of Maximum Likelihood Estimates*

| *Parameter* | | *DF* | *Estimate* | *Standard Error* | *Wald Chi-Square* | *Pr > ChiSq* |
|---|---|---|---|---|---|---|
| *Intercept* | | 1 | -1.7601 | 0.2132 | 68.1848 | <.0001 |
| *SEX* | *Male* | 1 | 0.4151 | 0.0672 | 38.1780 | <.0001 |
| *AGE* | *15 to 18 years* | 1 | 1.4579 | 0.1773 | 67.6083 | <.0001 |
| ***AGE*** | ***20 to 24 years*** | **1** | **0.1350** | **0.1782** | **0.5745** | **0.4485** |
| ***AGE*** | ***35 to 59 years*** | **1** | **0.0734** | **0.0977** | **0.5654** | **0.4521** |
| ***AGE*** | ***60 years and over*** | **1** | **0.2009** | **0.1234** | **2.6518** | **0.1034** |
| *CHILD* | *Child* | 1 | -0.1659 | 0.0776 | 4.5722 | 0.0325 |
| *STATUS* | *Not in labor force* | 1 | 0.3314 | 0.0880 | 14.1847 | 0.0002 |
| *STATUS* | *Unemployed* | 1 | 0.3761 | 0.1301 | 8.3551 | 0.0038 |
| *DEGREE* | *Advanced degree* | 1 | 0.8380 | 0.1469 | 32.5218 | <.0001 |
| *DEGREE* | *Bachelors degree* | 1 | 0.7317 | 0.1352 | 29.2883 | <.0001 |
| *DEGREE* | *High School Diploma* | 1 | 0.2655 | 0.1230 | 4.6609 | 0.0309 |
| *TTV* | | 1 | -0.00123 | 0.000240 | 26.0760 | <.0001 |
| *TSLEEP* | | 1 | -0.00083 | 0.000270 | 9.4614 | 0.0021 |

Taking the exponential of the coefficients of the categorical variables leads to the odds ratio, which is a measure of the effect size indicating the strength of association between the binary values. The odds ratio is a ratio of probabilities, which implies how much more (or less) likely is the participation in sports is under different set of conditions. For example, the odds ratio for sex (Male vs Female) is 1.514, which means that males are more likely than females to participate of sports. Table 22 shows the odds ratio estimates along with the 95% confidence limits. The coefficients highlighted in red indicate insignificant effect since the confidence limits include the odds ratio of 1 (e.g. equally likely). The only odds ratio less than 1 is the child, which implies that participation in sports is likely if the person has a child (reference level is no child).

**Table 22. Odds Ratio Estimates and Wald Confidence Intervals**

| *Label* | *Estimate* | *95% Confidence Limits* | |
|---|---|---|---|
| SEX Male vs Female | 1.514 | 1.328 | 1.728 |
| AGE 15 to 18 years vs 25 to 34 years | 4.297 | 3.036 | 6.083 |
| **AGE 20 to 24 years vs 25 to 34 years** | **1.145** | **0.807** | **1.623** |
| **AGE 35 to 59 years vs 25 to 34 years** | **1.076** | **0.889** | **1.303** |
| **AGE 60 years and over vs 25 to 34 years** | **1.223** | **0.960** | **1.557** |
| CHILD Child vs No Child | 0.847 | 0.728 | 0.986 |
| STATUS Not in labor force vs Employed | 1.393 | 1.172 | 1.655 |
| STATUS Unemployed vs Employed | 1.457 | 1.129 | 1.880 |
| DEGREE Advanced degree vs No High School Diploma | 2.312 | 1.733 | 3.083 |
| DEGREE Bachelors degree vs No High School Diploma | 2.079 | 1.595 | 2.709 |
| DEGREE High School Diploma vs No High School Diploma | 1.304 | 1.025 | 1.660 |

The results of the logistic model make sense. Males, teens, and people with no children are more likely to participate in sports than females, adults, and people with children respectively. The analysis also suggests that being people that are unemployed or not in the labor force are more likely to participate compared to people that are employed. This also makes sense since people that are employed have less free time than people that are not currently working. An interesting result that is not intuitive is that the higher the educational attainment, the more likely the person is to engage in sports activities. A graphical representation of the odds ratio comparison all pairs of variables (not against the reference level) is shown in Figure 20.

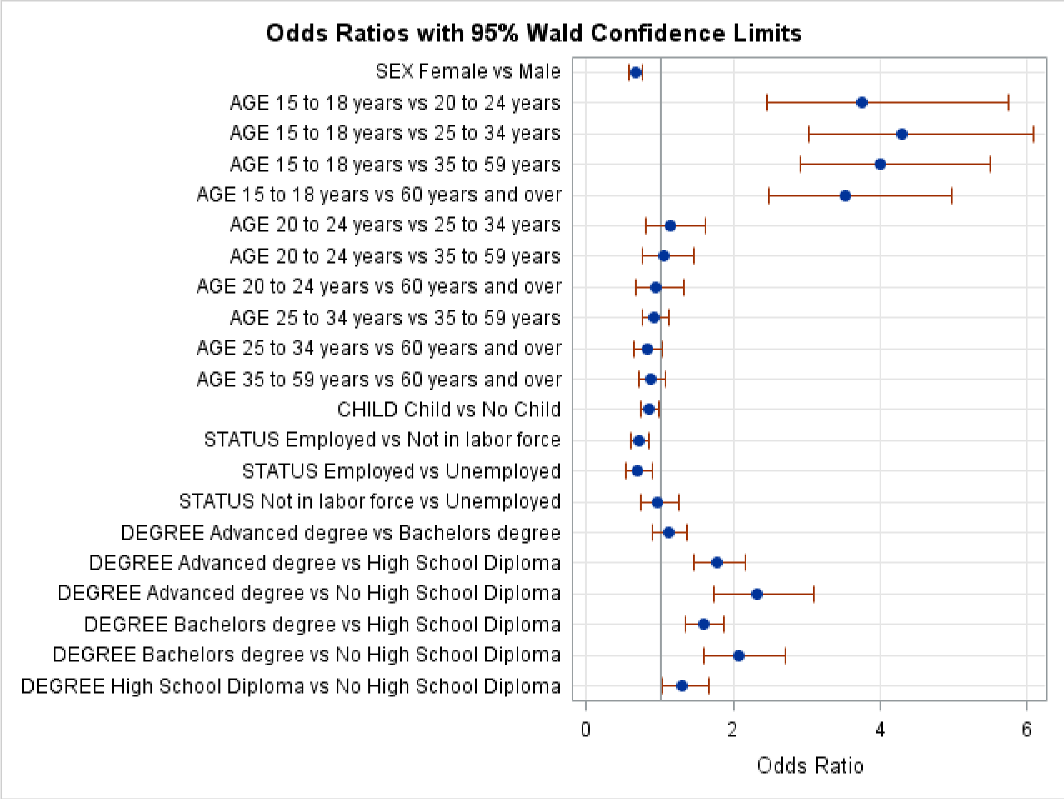



Figure 20. Odds Ratios for all pair of variables

The predicted probability for the participation in sports was plotted at specified levels of the categorical variables. The plots illustrate the conclusions discussed above, in which the probability decreases with increase watching TV and sleeping time. From Figure 21, it can be seen that the probability is higher for males and for people with no children.

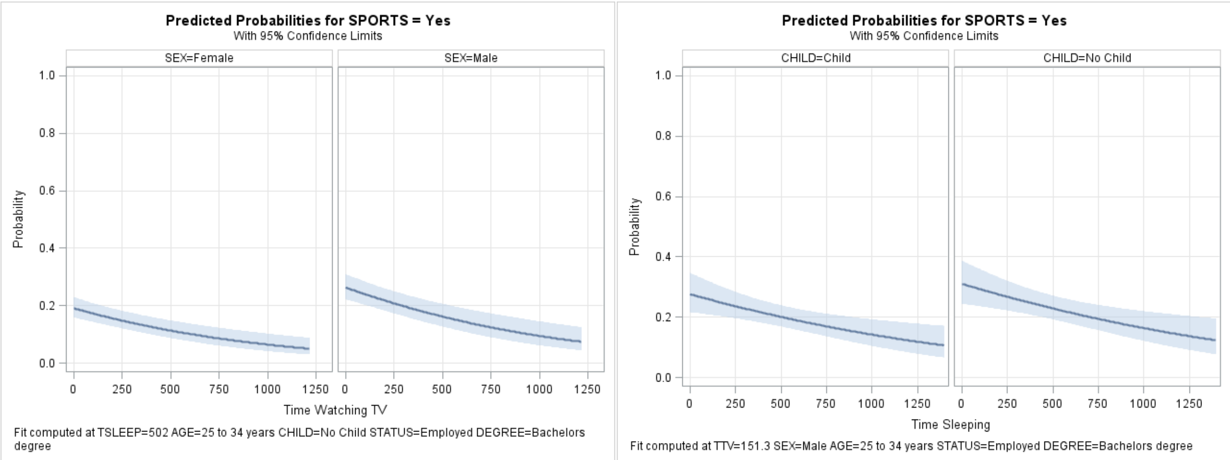



Figure 21. Predicted probabilities for levels of sex and child

The plots in Figure 22 indicate the same trend for time watching TV and sleeping holds. As expected from the previous analysis, it can be also seen that the teens, holders of advanced degrees and unemployed people have the highest probability in their corresponding groups.

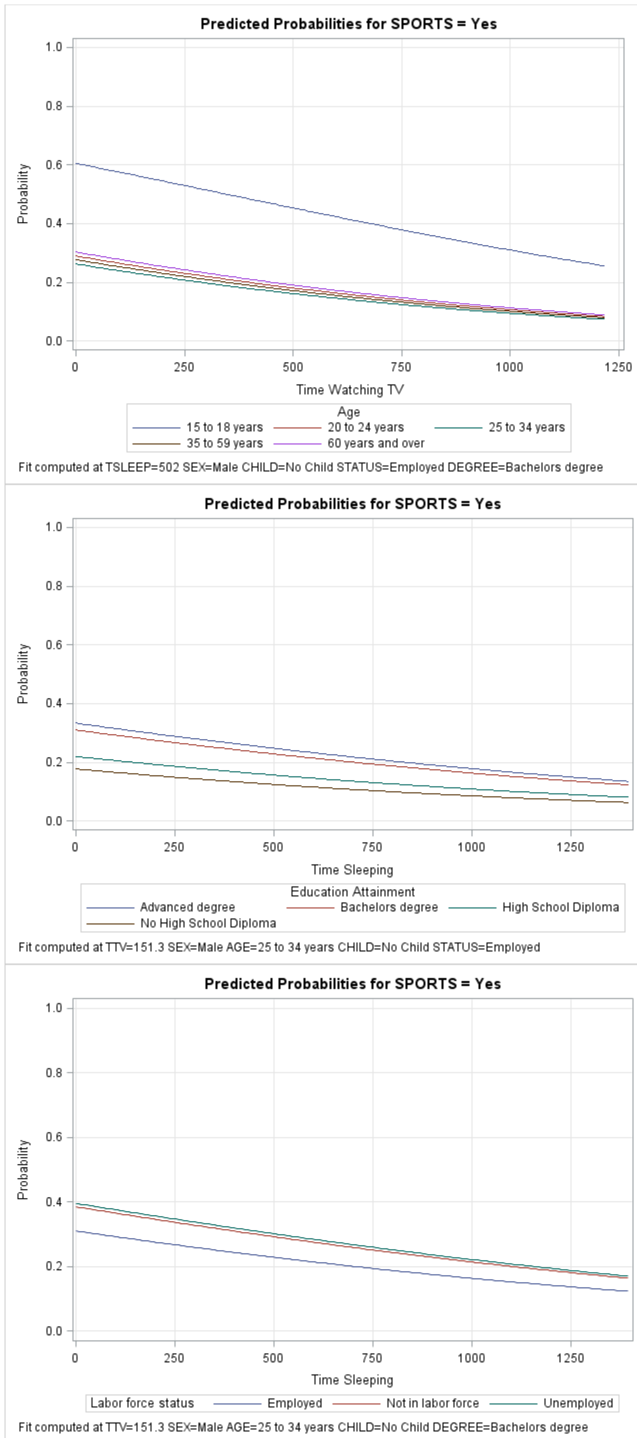



**Figure 22. Predicted probability for levels of age, degree and labor force status**

# 4. CONCLUSIONS

## 4.1 Summary

The analysis techniques to answer the research questions included an exploratory, associative, inferential, and predictive analysis. The inferences and conclusions based on the analysis are discussed in detail the results section and the main highlights are summarized below.

The exploratory analysis studied the distribution of the time spent on sports by the different categorical variables and revealed possible differences in time spent on sports and number of participants among the different groups of the categorical variables. The associative analysis showed a statistical significant but very weak negative correlation between the sports time and other time activities. Similar results were obtained for the participation in sports and other demographic variables. The associative analysis also showed signification association between sports participation and demographic variables. The inferential analysis was also able to detect significant differences in the median time spent on sports among some groups of the demographic variables. The predictive analysis fitted the best model and determined factors that had a significant effect in the probability of participating in sports. The odds ratio provided a measure to compare different groups and their likelihood in engaging in sports.

The results revealed relationship among variables and showed that the participation in sports and time spent in sports differs among different groups of the population. Furthermore, the results of the predictive model determined significant factors and seemed to be consistent with common expectations. In conclusion, the analysis techniques conducted in this report were successful in addressing the research questions related to the participation in sports activities.

## 4.2 Future Work

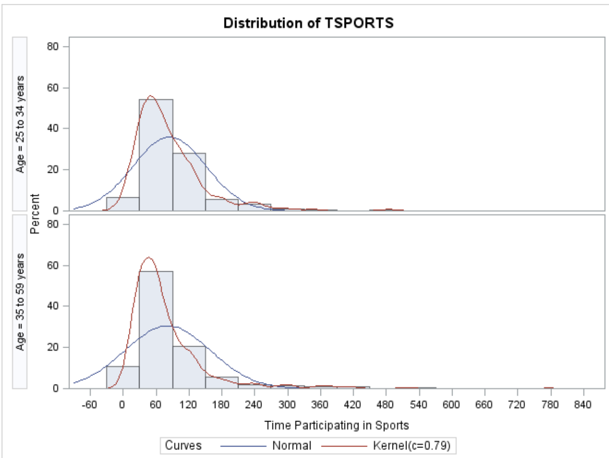
Future work might involve considering possible interactions to improve the predictive power of the model. In addition, a test data set should be used to test the accuracy of the model. An interesting follow up might involve comparing models for the weekdays and weekends, or using the weekday model and predict the participation of sports during weekends. Other statistical methods that can be applied include clustering, principal component analysis or factor analysis to reduce dimension or determine underlying features of people that are more likely to participate in sports. Finally, it might also be interesting to see if the predictive model depends on the type of sport.

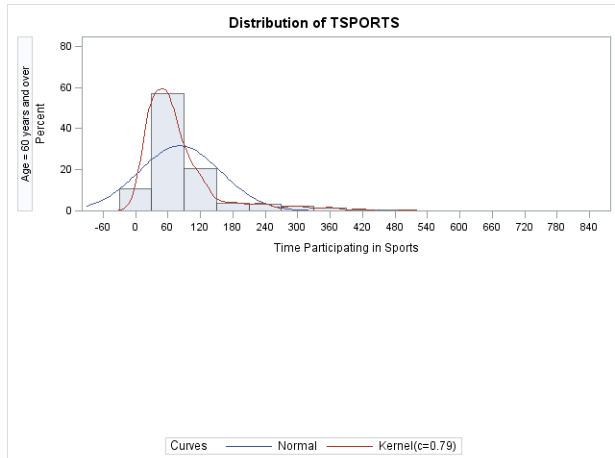
# REFERENCES

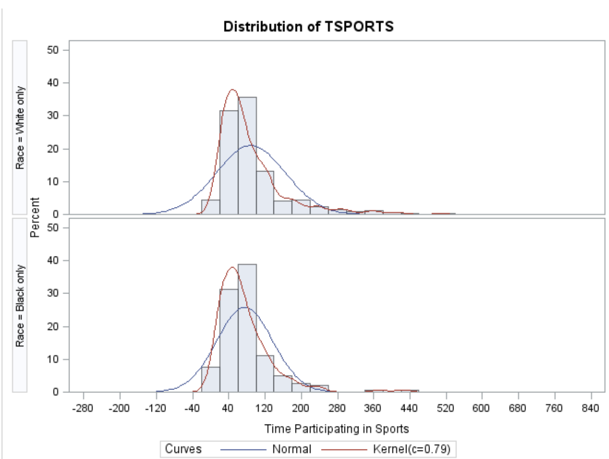
Der, Geoff, Everitt, Brian. *“A handbook of statistical analyses using SAS “,* Chapman and Hall/CRC, 3rd edition. 2009

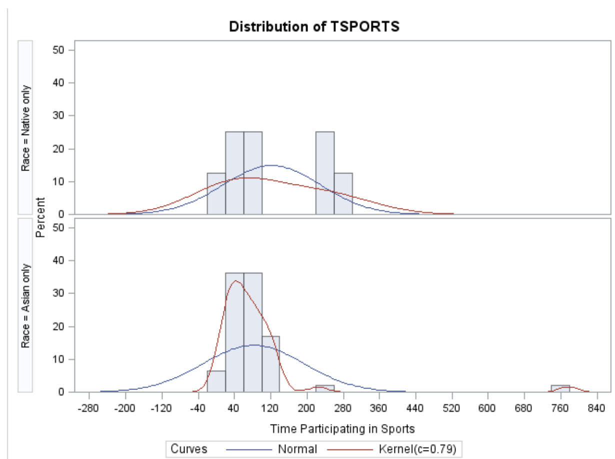
# APPENDIX

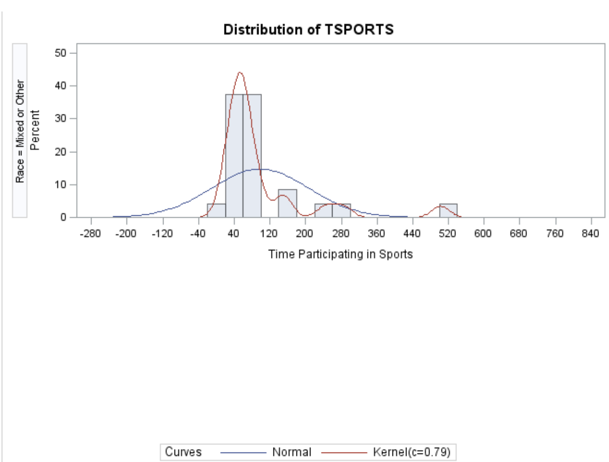
## A. Histograms of TSPORTS by categorical variable

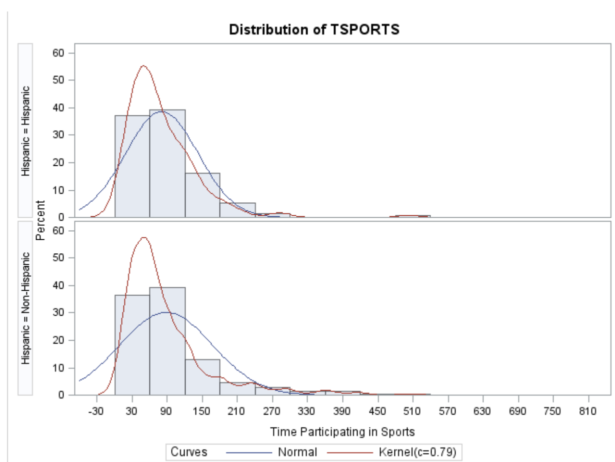


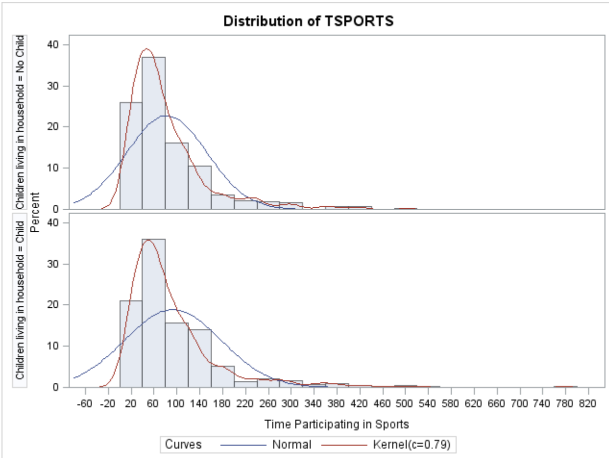



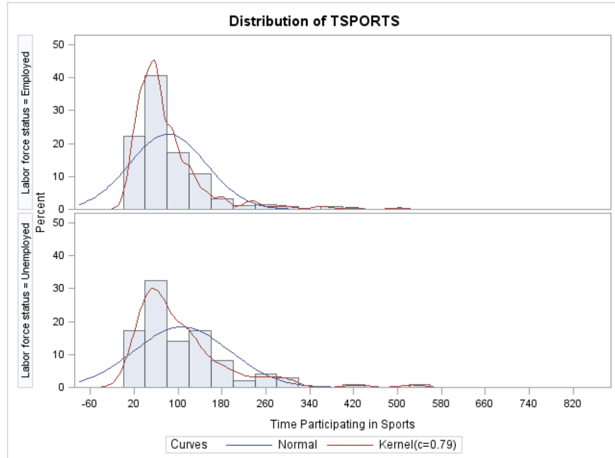



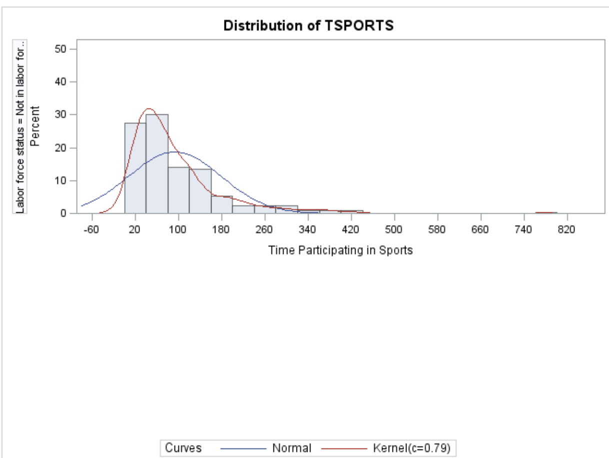



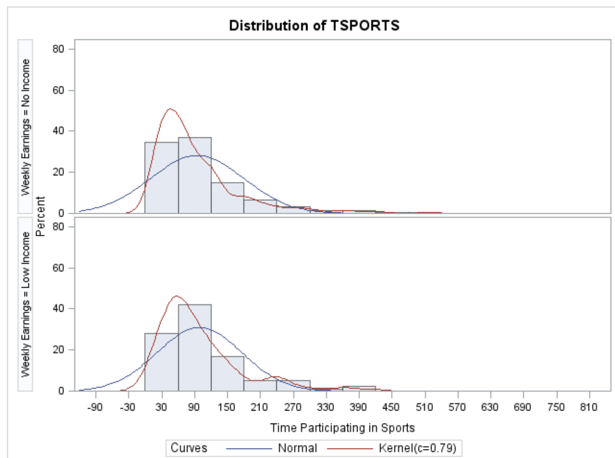



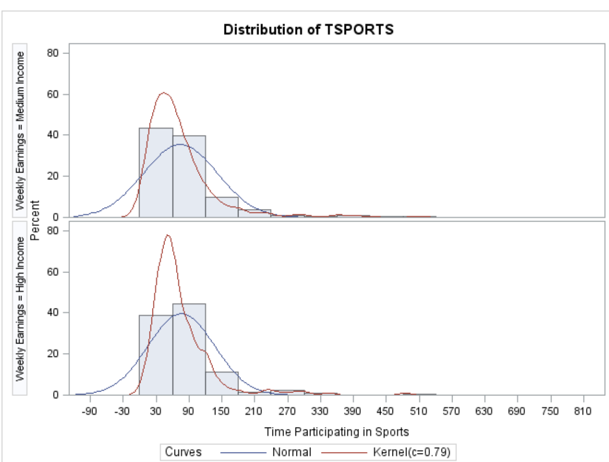



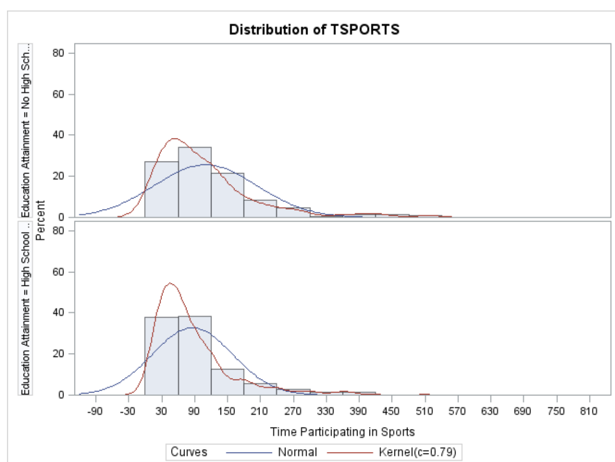

Distribution of TSPORTS
Children living in household = No Child
Children living in household = Child
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

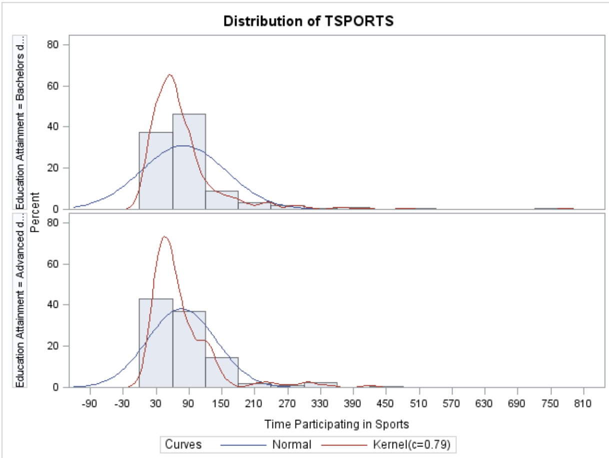

Distribution of TSPORTS
Labor force status = Employed
Labor force status = Unemployed
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

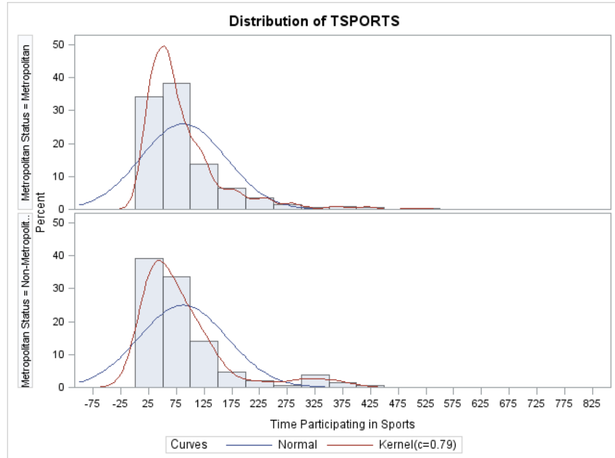

Distribution of TSPORTS
Labor force status = Not in labor for...
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

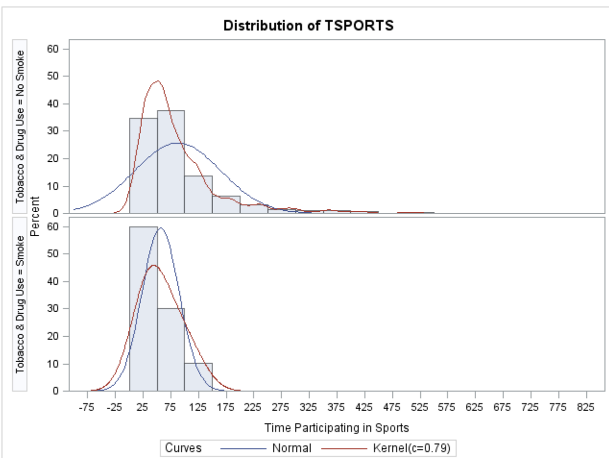

Distribution of TSPORTS
Weekly Earnings = No Income
Weekly Earnings = Low Income
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

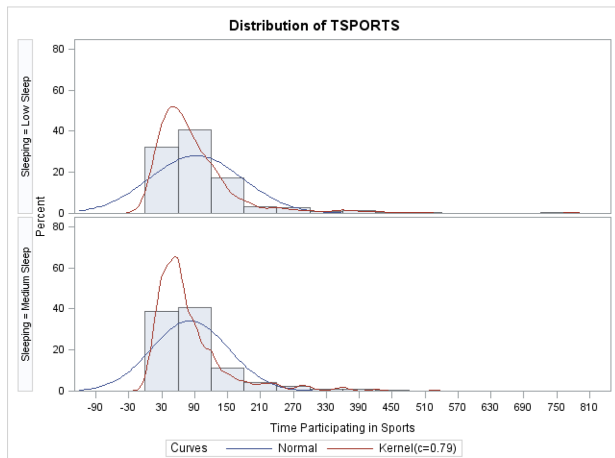

Distribution of TSPORTS
Weekly Earnings = Medium Income
Weekly Earnings = High Income
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

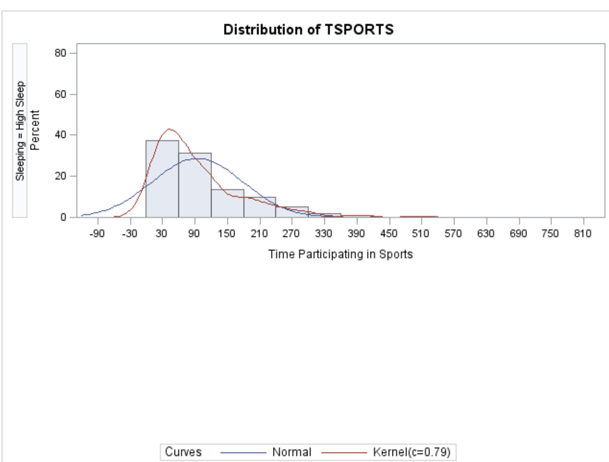

Distribution of TSPORTS
Education Attainment = No High Sch...
Education Attainment = High School ...
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

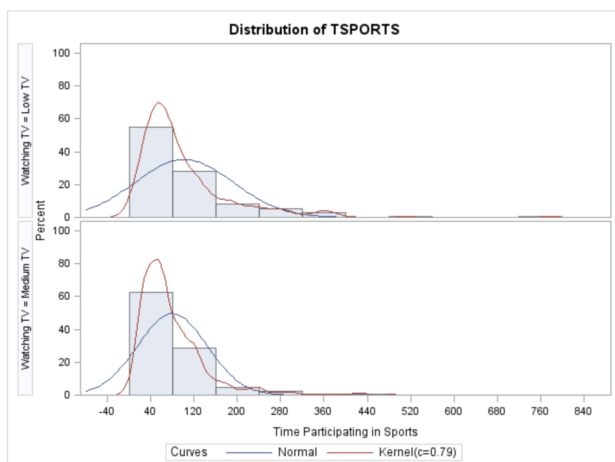

Distribution of TSPORTS
Education Attainment = Bachelors d...
Education Attainment = Advanced d...
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

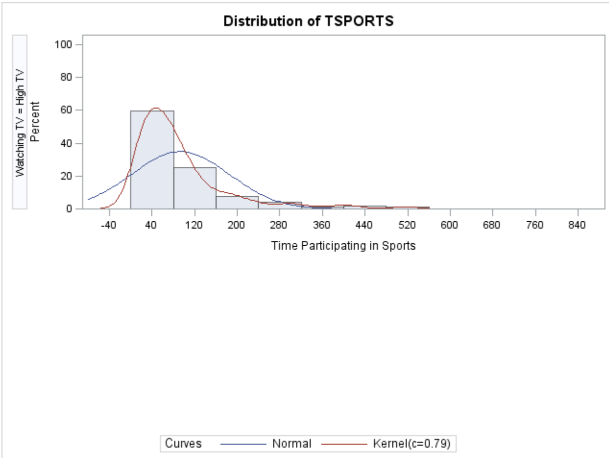

Distribution of TSPORTS
Metropolitan Status = Metropolitan
Metropolitan Status = Non-Metropolit...
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

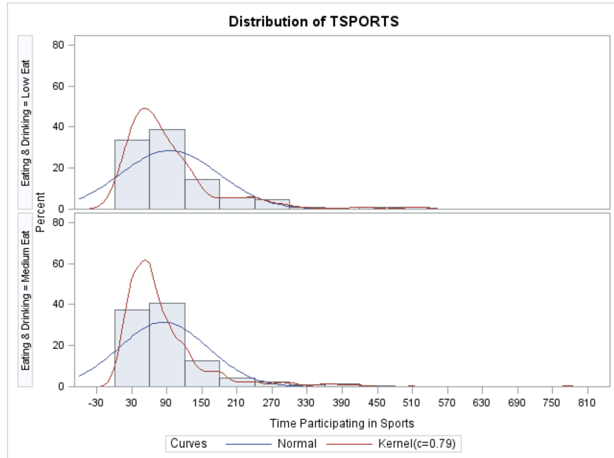

Distribution of TSPORTS
Tobacco & Drug Use = No Smoke
Tobacco & Drug Use = Smoke
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

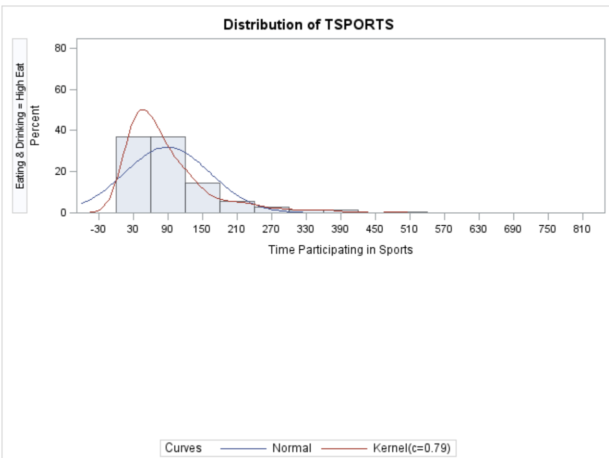

Distribution of TSPORTS
Sleeping = Low Sleep
Sleeping = Medium Sleep
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)


Distribution of TSPORTS
Sleeping = High Sleep
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)


Distribution of TSPORTS
Watching TV = Low TV
Watching TV = Medium TV
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)


Distribution of TSPORTS
Watching TV = High TV
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)


Distribution of TSPORTS
Eating & Drinking = Low Eat
Eating & Drinking = Medium Eat
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)


Distribution of TSPORTS
Eating & Drinking = High Eat
Percent
Time Participating in Sports
Curves
Normal
Kernel(c=0.79)

## B. Chi-square tests for sports and categorical variables

Table of SPORTS by AGE

| SPORTS(Participating in Sports) Frequency Col Pct | AGE(Age) 15 to 18 years | 20 to 24 years | 25 to 34 years | 35 to 59 years | 60 years and over | Total |
|---|---|---|---|---|---|---|
| No | 228<br>63.69 | 270<br>84.11 | 913<br>84.15 | 2542<br>82.96 | 1351<br>81.19 | 5304 |
| Yes | 130<br>36.31 | 51<br>15.89 | 172<br>15.85 | 522<br>17.04 | 313<br>18.81 | 1188 |
| Total | 358 | 321 | 1085 | 3064 | 1664 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 4 | 86.8509 | <.0001 |
| Likelihood Ratio Chi-Square | 4 | 74.0175 | <.0001 |
| Mantel-Haenszel Chi-Square | 1 | 2.3954 | 0.1217 |
| Phi Coefficient | | 0.1157 | |
| Contingency Coefficient | | 0.1149 | |
| Cramer's V | | 0.1157 | |

Table of SPORTS by RACE

| SPORTS(Participating in Sports) Frequency Col Pct | RACE(Race) White only | Black only | Native only | Asian only | Mixed or Other | Total |
|---|---|---|---|---|---|---|
| No | 4162<br>81.18 | 815<br>84.98 | 54<br>87.10 | 182<br>79.48 | 91<br>79.13 | 5304 |
| Yes | 965<br>18.82 | 144<br>15.02 | 8<br>12.90 | 47<br>20.52 | 24<br>20.87 | 1188 |
| Total | 5127 | 959 | 62 | 229 | 115 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 4 | 10.3267 | 0.0353 |
| Likelihood Ratio Chi-Square | 4 | 10.7454 | 0.0296 |
| Mantel-Haenszel Chi-Square | 1 | 0.1890 | 0.6637 |
| Phi Coefficient | | 0.0399 | |
| Contingency Coefficient | | 0.0399 | |
| Cramer's V | | 0.0399 | |

Table of SPORTS by HISPANIC

| SPORTS(Participating in Sports) Frequency Col Pct | HISPANIC(Hispanic) Hispanic | Non-Hispanic | Total |
|---|---|---|---|
| No | 764<br>83.77 | 4540<br>81.36 | 5304 |
| Yes | 148<br>16.23 | 1040<br>18.64 | 1188 |
| Total | 912 | 5580 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 1 | 3.0450 | 0.0810 |
| Likelihood Ratio Chi-Square | 1 | 3.1246 | 0.0771 |
| Continuity Adj. Chi-Square | 1 | 2.8860 | 0.0894 |
| Mantel-Haenszel Chi-Square | 1 | 3.0446 | 0.0810 |
| Phi Coefficient | | 0.0217 | |
| Contingency Coefficient | | 0.0217 | |
| Cramer's V | | 0.0217 | |

*Table of SPORTS by CHILD*

| SPORTS(Participating in Sports) | CHILD(Children living in household) | | |
|---|---|---|---|
| Frequency<br>Col Pct | No Child | Child | Total |
| No | 2814<br>81.42 | 2490<br>82.02 | 5304 |
| Yes | 642<br>18.58 | 546<br>17.98 | 1188 |
| Total | 3456 | 3036 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 1 | 0.3791 | 0.5381 |
| Likelihood Ratio Chi-Square | 1 | 0.3793 | 0.5380 |
| Continuity Adj. Chi-Square | 1 | 0.3405 | 0.5595 |
| Mantel-Haenszel Chi-Square | 1 | 0.3791 | 0.5381 |
| Phi Coefficient | | -0.0076 | |
| Contingency Coefficient | | 0.0076 | |
| Cramer's V | | -0.0076 | |

*Table of SPORTS by STATUS*

| SPORTS(Participating in Sports) | STATUS(Labor force status) | | | |
|---|---|---|---|---|
| Frequency<br>Col Pct | Employed | Unemployed | Not in labor force | Total |
| No | 3257<br>82.62 | 353<br>78.10 | 1694<br>80.74 | 5304 |
| Yes | 685<br>17.38 | 99<br>21.90 | 404<br>19.26 | 1188 |
| Total | 3942 | 452 | 2098 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 2 | 7.4540 | 0.0241 |
| Likelihood Ratio Chi-Square | 2 | 7.2860 | 0.0262 |
| Mantel-Haenszel Chi-Square | 1 | 3.8622 | 0.0494 |
| Phi Coefficient | | 0.0339 | |
| Contingency Coefficient | | 0.0339 | |
| Cramer's V | | 0.0339 | |

Table of SPORTS by EARNINGS

| SPORTS(Participating in Sports) | EARNINGS(Weekly Earnings) | | | | |
|---|---|---|---|---|---|
| Frequency<br>Col Pct | No Income | Low Income | Medium Income | High Income | Total |
| No | 2454 | 828 | 1402 | 620 | 5304 |
| | 80.78 | 85.89 | 83.30 | 76.83 | |
| Yes | 584 | 136 | 281 | 187 | 1188 |
| | 19.22 | 14.11 | 16.70 | 23.17 | |
| Total | 3038 | 964 | 1683 | 807 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 3 | 28.7715 | <.0001 |
| Likelihood Ratio Chi-Square | 3 | 28.8100 | <.0001 |
| Mantel-Haenszel Chi-Square | 1 | 7.8671 | 0.0050 |
| Phi Coefficient | | 0.0666 | |
| Contingency Coefficient | | 0.0664 | |
| Cramer's V | | 0.0666 | |

Table of SPORTS by DEGREE

| SPORTS(Participating in Sports) | DEGREE(Education Attainment) | | | | |
|---|---|---|---|---|---|
| Frequency<br>Col Pct | No High School Diploma | High School Diploma | Bachelors degree | Advanced degree | Total |
| No | 853 | 2872 | 1025 | 554 | 5304 |
| | 80.55 | 84.82 | 78.01 | 75.58 | |
| Yes | 206 | 514 | 289 | 179 | 1188 |
| | 19.45 | 15.18 | 21.99 | 24.42 | |
| Total | 1059 | 3386 | 1314 | 733 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 3 | 53.3411 | <.0001 |
| Likelihood Ratio Chi-Square | 3 | 52.4778 | <.0001 |
| Mantel-Haenszel Chi-Square | 1 | 6.4279 | 0.0112 |
| Phi Coefficient | | 0.0906 | |
| Contingency Coefficient | | 0.0903 | |
| Cramer's V | | 0.0906 | |

Table of SPORTS by METRO

| SPORTS(Participating in Sports) | METRO(Metropolitan Status) | | |
|---|---|---|---|
| Frequency<br>Col Pct | Metropolitan | Non-Metropolitan | Total |
| No | 4358 | 946 | 5304 |
| | 81.32 | 83.50 | |
| Yes | 1001 | 187 | 1188 |
| | 18.68 | 16.50 | |
| Total | 5359 | 1133 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 1 | 2.9566 | 0.0855 |
| Likelihood Ratio Chi-Square | 1 | 3.0192 | 0.0823 |
| Continuity Adj. Chi-Square | 1 | 2.8130 | 0.0935 |
| Mantel-Haenszel Chi-Square | 1 | 2.9561 | 0.0856 |
| Phi Coefficient | | -0.0213 | |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Contingency Coefficient | | 0.0213 | |
| Cramer's V | | -0.0213 | |

| Table of SPORTS by SMOKE | | | |
|---|---|---|---|
| SPORTS(Participating in Sports) | SMOKE(Tobacco & Drug Use) | | |
| Frequency<br>Col Pct | No Smoke | Smoke | Total |
| No | 5207<br>81.55 | 97<br>90.65 | 5304 |
| Yes | 1178<br>18.45 | 10<br>9.35 | 1188 |
| Total | 6385 | 107 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 1 | 5.8336 | 0.0157 |
| Likelihood Ratio Chi-Square | 1 | 6.8314 | 0.0090 |
| Continuity Adj. Chi-Square | 1 | 5.2406 | 0.0221 |
| Mantel-Haenszel Chi-Square | 1 | 5.8327 | 0.0157 |
| Phi Coefficient | | -0.0300 | |
| Contingency Coefficient | | 0.0300 | |
| Cramer's V | | -0.0300 | |

| Table of SPORTS by SLEEP | | | | |
|---|---|---|---|---|
| SPORTS(Participating in Sports) | SLEEP(Sleeping) | | | |
| Frequency<br>Col Pct | Low Sleep | Medium Sleep | High Sleep | Total |
| No | 1597<br>81.02 | 2646<br>80.67 | 1061<br>85.50 | 5304 |
| Yes | 374<br>18.98 | 634<br>19.33 | 180<br>14.50 | 1188 |
| Total | 1971 | 3280 | 1241 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 2 | 14.8832 | 0.0006 |
| Likelihood Ratio Chi-Square | 2 | 15.5669 | 0.0004 |
| Mantel-Haenszel Chi-Square | 1 | 3.2448 | 0.0717 |
| Phi Coefficient | | 0.0479 | |
| Contingency Coefficient | | 0.0478 | |
| Cramer's V | | 0.0479 | |

| Table of SPORTS by TV | | | | |
|---|---|---|---|---|
| SPORTS(Participating in Sports) | TV(Watching TV) | | | |
| Frequency<br>Col Pct | Low TV | Medium TV | High TV | Total |
| No | 1259<br>80.34 | 2911<br>80.86 | 1134<br>85.58 | 5304 |
| Yes | 308<br>19.66 | 689<br>19.14 | 191<br>14.42 | 1188 |
| Total | 1567 | 3600 | 1325 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 2 | 16.9956 | 0.0002 |
| Likelihood Ratio Chi-Square | 2 | 17.7698 | 0.0001 |
| Mantel-Haenszel Chi-Square | 1 | 16.9796 | <.0001 |
| Phi Coefficient | | 0.0512 | |
| Contingency Coefficient | | 0.0511 | |
| Cramer's V | | 0.0512 | |

Table of SPORTS by EAT

| SPORTS(Participating in Sports) | EAT(Eating & Drinking) | | | |
|---|---|---|---|---|
| Frequency Col Pct | Low Eat | Medium Eat | High Eat | Total |
| No | 1110 | 2894 | 1300 | 5304 |
| | 84.15 | 81.43 | 80.30 | |
| Yes | 209 | 660 | 319 | 1188 |
| | 15.85 | 18.57 | 19.70 | |
| Total | 1319 | 3554 | 1619 | 6492 |

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 2 | 7.6230 | 0.0221 |
| Likelihood Ratio Chi-Square | 2 | 7.7803 | 0.0204 |
| Mantel-Haenszel Chi-Square | 1 | 5.6053 | 0.0179 |
| Phi Coefficient | | 0.0343 | |
| Contingency Coefficient | | 0.0342 | |
| Cramer's V | | 0.0343 | |

## C. Kruskal-Wallis Test for TSPORTS and categorical variables

Order: Age, Race, Hispanic, Child, Status, Earnings, Degree, Metro, Smoke, Sleep, TV

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 57.1979 |
| *DF* | 4 |
| *Pr > Chi-Square* | <.0001 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 5.3528 |
| *DF* | 4 |
| *Pr > Chi-Square* | 0.2530 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 0.0140 |
| *DF* | 1 |
| *Pr > Chi-Square* | 0.9058 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 4.6337 |
| *DF* | 1 |
| *Pr > Chi-Square* | 0.0313 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 11.2779 |
| *DF* | 2 |
| *Pr > Chi-Square* | 0.0036 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 21.5592 |
| *DF* | 3 |
| *Pr > Chi-Square* | <.0001 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 20.5559 |
| *DF* | 3 |
| *Pr > Chi-Square* | 0.0001 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 0.9693 |
| *DF* | 1 |
| *Pr > Chi-Square* | 0.3248 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 1.7498 |
| *DF* | 1 |
| *Pr > Chi-Square* | 0.1859 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 5.5300 |
| *DF* | 2 |
| *Pr > Chi-Square* | 0.0630 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 13.2072 |
| *DF* | 2 |
| *Pr > Chi-Square* | 0.0014 |

| Kruskal-Wallis Test | |
|---|---|
| *Chi-Square* | 3.4064 |
| *DF* | 2 |
| *Pr > Chi-Square* | 0.1821 |

## D. Logistic Regression Stepwise

| Model Information | | |
|---|---|---|
| *Data Set* | WORK.ATUSCAT | |
| *Response Variable* | SPORTS | Participating in Sports |
| *Number of Response Levels* | 2 | |
| *Model* | binary logit | |
| *Optimization Technique* | Fisher's scoring | |

| | |
|---|---|
| *Number of Observations Read* | 6492 |
| *Number of Observations Used* | 6492 |

| Response Profile | | |
|---|---|---|
| *Ordered Value* | *SPORTS* | *Total Frequency* |
| *1* | Yes | 1188 |
| *2* | No | 5304 |

***Probability modeled is SPORTS='Yes'.***

***Stepwise Selection Procedure***

| *Class* | *Value* | *Design Variables* | | | |
|---|---|---|---|---|---|
| *SEX* | *Female* | 0 | | | |
| | *Male* | 1 | | | |
| *AGE* | *15 to 18 years* | 1 | 0 | 0 | 0 |
| | *20 to 24 years* | 0 | 1 | 0 | 0 |
| | *25 to 34 years* | 0 | 0 | 0 | 0 |
| | *35 to 59 years* | 0 | 0 | 1 | 0 |
| | *60 years and over* | 0 | 0 | 0 | 1 |
| *RACE* | *Asian only* | 1 | 0 | 0 | 0 |
| | *Black only* | 0 | 1 | 0 | 0 |
| | *Mixed or Other* | 0 | 0 | 1 | 0 |
| | *Native only* | 0 | 0 | 0 | 1 |
| | *White only* | 0 | 0 | 0 | 0 |
| *HISPANIC* | *Hispanic* | 1 | | | |
| | *Non-Hispanic* | 0 | | | |
| *CHILD* | *Child* | 1 | | | |
| | *No Child* | 0 | | | |
| *STATUS* | *Employed* | 0 | 0 | | |
| | *Not in labor force* | 1 | 0 | | |
| | *Unemployed* | 0 | 1 | | |
| *EARNINGS* | *High Income* | 1 | 0 | 0 | |
| | *Low Income* | 0 | 0 | 0 | |
| | *Medium Income* | 0 | 1 | 0 | |
| | *No Income* | 0 | 0 | 1 | |

*Class Level Information*

| Class Level Information | | | | |
|---|---|---|---|---|
| | | Design Variables | | |
| Class | Value | | | |
| DEGREE | Advanced degree | 1 | 0 | 0 |
| | Bachelors degree | 0 | 1 | 0 |
| | High School Diploma | 0 | 0 | 1 |
| | No High School Diploma | 0 | 0 | 0 |
| METRO | Metropolitan | 1 | | |
| | Non-Metropolitan | 0 | | |
| SMOKE | No Smoke | 0 | | |
| | Smoke | 1 | | |

***Step 0. Intercept entered:***

| Model Convergence Status |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| | | |
|---|---|---|
| -2 Log L | = | 6179.136 |

| Residual Chi-Square Test | | |
|---|---|---|
| Chi-Square | DF | Pr > ChiSq |
| 252.5388 | 25 | <.0001 |

***Step 1. Effect AGE entered:***

| Model Convergence Status |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| AIC | 6181.136 | 6115.118 |
| SC | 6187.914 | 6149.010 |
| -2 Log L | 6179.136 | 6105.118 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 74.0175 | 4 | <.0001 |
| Score | 86.8509 | 4 | <.0001 |
| Wald | 81.1549 | 4 | <.0001 |

| Residual Chi-Square Test | | |
|---|---|---|
| Chi-Square | DF | Pr > ChiSq |
| 169.8726 | 21 | <.0001 |

***Note:*** `No effects for the model in Step 1 are removed.`

***Step 2. Effect DEGREE entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

*Model Fit Statistics*

| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
|---|---|---|
| *AIC* | 6181.136 | 6043.900 |
| *SC* | 6187.914 | 6098.127 |
| *-2 Log L* | 6179.136 | 6027.900 |

*Testing Global Null Hypothesis: BETA=0*

| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|---|
| *Likelihood Ratio* | 151.2358 | 7 | <.0001 |
| *Score* | 162.1985 | 7 | <.0001 |
| *Wald* | 154.0215 | 7 | <.0001 |

*Residual Chi-Square Test*

| *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|
| 93.4175 | 18 | <.0001 |

***Note:*** No effects for the model in Step 2 are removed.

***Step 3. Effect SEX entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

*Model Fit Statistics*

| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
|---|---|---|
| *AIC* | 6181.136 | 6013.096 |
| *SC* | 6187.914 | 6074.101 |
| *-2 Log L* | 6179.136 | 5995.096 |

*Testing Global Null Hypothesis: BETA=0*

| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|---|
| *Likelihood Ratio* | 184.0395 | 8 | <.0001 |
| *Score* | 194.6546 | 8 | <.0001 |
| *Wald* | 184.3872 | 8 | <.0001 |

*Residual Chi-Square Test*

| *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|
| 60.5262 | 17 | <.0001 |

**Note:** No effects for the model in Step 3 are removed.

***Step 4. Effect TTV entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

*Model Fit Statistics*

| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
|---|---|---|
| *AIC* | 6181.136 | 5995.222 |
| *SC* | 6187.914 | 6063.005 |
| *-2 Log L* | 6179.136 | 5975.222 |

*Testing Global Null Hypothesis: BETA=0*

| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|---|
| *Likelihood Ratio* | 203.9144 | 9 | <.0001 |
| *Score* | 212.9865 | 9 | <.0001 |
| *Wald* | 201.2523 | 9 | <.0001 |

*Residual Chi-Square Test*

| *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|
| 41.6771 | 16 | 0.0004 |

**Note:** No effects for the model in Step 4 are removed.

***Step 5. Effect STATUS entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

*Model Fit Statistics*

| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
|---|---|---|
| *AIC* | 6181.136 | 5985.708 |
| *SC* | 6187.914 | 6067.048 |
| *-2 Log L* | 6179.136 | 5961.708 |

*Testing Global Null Hypothesis: BETA=0*

| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
|---|---|---|---|
| *Likelihood Ratio* | 217.4283 | 11 | <.0001 |
| *Score* | 226.2065 | 11 | <.0001 |
| *Wald* | 213.2254 | 11 | <.0001 |

| Residual Chi-Square Test | | |
|---|---|---|
| *Chi-Square* | *DF* | *Pr > ChiSq* |
| 28.3661 | 14 | 0.0127 |

***Note:*** No effects for the model in Step 5 are removed.

***Step 6. Effect TSLEEP entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| Model Fit Statistics | | |
|---|---|---|
| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
| *AIC* | 6181.136 | 5978.638 |
| *SC* | 6187.914 | 6066.756 |
| *-2 Log L* | 6179.136 | 5952.638 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| *Test* | *Chi-Square* | *DF* | *Pr > ChiSq* |
| *Likelihood Ratio* | 226.4982 | 12 | <.0001 |
| *Score* | 235.1340 | 12 | <.0001 |
| *Wald* | 221.3069 | 12 | <.0001 |

| Residual Chi-Square Test | | |
|---|---|---|
| *Chi-Square* | *DF* | *Pr > ChiSq* |
| 19.4464 | 13 | 0.1099 |

***Note:*** No effects for the model in Step 6 are removed.

***Step 7. Effect CHILD entered:***

| *Model Convergence Status* |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| Model Fit Statistics | | |
|---|---|---|
| *Criterion* | *Intercept Only* | *Intercept and Covariates* |
| *AIC* | 6181.136 | 5976.069 |
| *SC* | 6187.914 | 6070.965 |
| *-2 Log L* | 6179.136 | 5948.069 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 231.0671 | 13 | <.0001 |
| Score | 239.5551 | 13 | <.0001 |
| Wald | 225.3341 | 13 | <.0001 |

| Residual Chi-Square Test | | |
|---|---|---|
| Chi-Square | DF | Pr > ChiSq |
| 14.7921 | 12 | 0.2530 |

**Note:** No effects for the model in Step 7 are removed.

**Note:** No (additional) effects met the 0.05 significance level for entry into the model.

| Summary of Stepwise Selection | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Effect | | | Number | Score | Wald | | Variable |
| Step | Entered | Removed | DF | In | Chi-Square | Chi-Square | Pr > ChiSq | Label |
| 1 | AGE | | 4 | 1 | 86.8509 | | <.0001 | Age |
| 2 | DEGREE | | 3 | 2 | 78.7494 | | <.0001 | Education Attainment |
| 3 | SEX | | 1 | 3 | 32.9711 | | <.0001 | Sex |
| 4 | TTV | | 1 | 4 | 18.9077 | | <.0001 | Time Watching TV |
| 5 | STATUS | | 2 | 5 | 13.6808 | | 0.0011 | Labor force status |
| 6 | TSLEEP | | 1 | 6 | 8.9586 | | 0.0028 | Time Sleeping |
| 7 | CHILD | | 1 | 7 | 4.5772 | | 0.0324 | Children living in household |

| Type 3 Analysis of Effects | | | |
|---|---|---|---|
| | | Wald | |
| Effect | DF | Chi-Square | Pr > ChiSq |
| SEX | 1 | 38.1780 | <.0001 |
| AGE | 4 | 76.4926 | <.0001 |
| CHILD | 1 | 4.5722 | 0.0325 |
| STATUS | 2 | 18.0390 | 0.0001 |
| DEGREE | 3 | 61.3474 | <.0001 |
| TTV | 1 | 26.0760 | <.0001 |
| TSLEEP | 1 | 9.4614 | 0.0021 |

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Standard | Wald | |
| Parameter | | DF | Estimate | Error | Chi-Square | Pr > ChiSq |
| Intercept | | 1 | -1.7601 | 0.2132 | 68.1848 | <.0001 |
| SEX | Male | 1 | 0.4151 | 0.0672 | 38.1780 | <.0001 |
| AGE | 15 to 18 years | 1 | 1.4579 | 0.1773 | 67.6083 | <.0001 |
| AGE | 20 to 24 years | 1 | 0.1350 | 0.1782 | 0.5745 | 0.4485 |
| AGE | 35 to 59 years | 1 | 0.0734 | 0.0977 | 0.5654 | 0.4521 |
| AGE | 60 years and over | 1 | 0.2009 | 0.1234 | 2.6518 | 0.1034 |
| CHILD | Child | 1 | -0.1659 | 0.0776 | 4.5722 | 0.0325 |
| STATUS | Not in labor force | 1 | 0.3314 | 0.0880 | 14.1847 | 0.0002 |
| STATUS | Unemployed | 1 | 0.3761 | 0.1301 | 8.3551 | 0.0038 |
| DEGREE | Advanced degree | 1 | 0.8380 | 0.1469 | 32.5218 | <.0001 |
| DEGREE | Bachelors degree | 1 | 0.7317 | 0.1352 | 29.2883 | <.0001 |
| DEGREE | High School Diploma | 1 | 0.2655 | 0.1230 | 4.6609 | 0.0309 |
| TTV | | 1 | -0.00123 | 0.000240 | 26.0760 | <.0001 |
| TSLEEP | | 1 | -0.00083 | 0.000270 | 9.4614 | 0.0021 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| *Percent Concordant* | 63.3 | *Somers' D* | 0.265 |
| *Percent Discordant* | 36.7 | *Gamma* | 0.265 |
| *Percent Tied* | 0.0 | *Tau-a* | 0.079 |
| *Pairs* | 6301152 | *c* | 0.633 |

*Odds Ratio Estimates and Wald Confidence Intervals*

| *Label* | *Estimate* | *95% Confidence Limits* | |
|---|---|---|---|
| SEX Male vs Female | 1.514 | 1.328 | 1.728 |
| AGE 15 to 18 years vs 25 to 34 years | 4.297 | 3.036 | 6.083 |
| AGE 20 to 24 years vs 25 to 34 years | 1.145 | 0.807 | 1.623 |
| AGE 35 to 59 years vs 25 to 34 years | 1.076 | 0.889 | 1.303 |
| AGE 60 years and over vs 25 to 34 years | 1.223 | 0.960 | 1.557 |
| CHILD Child vs No Child | 0.847 | 0.728 | 0.986 |
| STATUS Not in labor force vs Employed | 1.393 | 1.172 | 1.655 |
| STATUS Unemployed vs Employed | 1.457 | 1.129 | 1.880 |
| DEGREE Advanced degree vs No High School Diploma | 2.312 | 1.733 | 3.083 |
| DEGREE Bachelors degree vs No High School Diploma | 2.079 | 1.595 | 2.709 |
| DEGREE High School Diploma vs No High School Diploma | 1.304 | 1.025 | 1.660 |