```sas
*/
**********************************************************************
***
* Class: STAT 448
*
* Author: Luis Lin
*
* Date: 12/12/12
*
* Title: American Time Use Survey
*
*
*
* Details:
*
* The file is divided in secionts for the format,input data,creating
data *
* sets for analysis, and the analysis.
*
**********************************************************************
****/;

ods html close;
ods preferences;
ods html newfile=proc;


*rtf;
ods html close;

* nodate and nonumber used to leave the date and page number
* off the resulting document;
options nodate nonumber;

* the following statement will just keep the 'The SAS System'
title from being genrated on each page of the report;
title ;

* nogtitle leaves titles off graphics;
* make sure the file's directory is one you can write to;
ods rtf file='C:\Stat 448\results.rtf' nogtitle style=journal;
*ods rtf file='f:\448\Homework 5\results.rtf' nogtitle style=journal;

* noproctitle supresses printing of the name of the procedure
* in the result file;
ods noproctitle;




**********************************************************************
*;
**********************************************************************
*;
```

```
* Create formats;
* Note: formats are different from the ones given in the codebook;
proc format;

   value WEEKDAYf  0='Non-weekday'
                   1='Weekday';

   value SEXf      1='Male'
                   2='Female';

   value RACEf     1='White only'
                   2='Black only'
                   3='Native only'
                   4='Asian only'
                   other='Mixed or Other';

   value HISPANICf 1='Hispanic'
                   2='Non-Hispanic';

   value STATUSf   1,2='Employed'
                   3,4='Unemployed'
                   5='Not in labor force';

   value DEGREEf   31-38='No High School Diploma'
                   39-42='High School Diploma'
                   43='Bachelors degree'
                   44-46='Advanced degree';

   value METROf    1='Metropolitan'
                    2,3='Non-Metropolitan';

   value AGEf      low-18='15 to 18 years'
                    19-24='20 to 24 years'
                   25-34='25 to 34 years'
                   35-59='35 to 59 years'
                   60-high='60 years and over';

   value CHILDf    0='No Child'
                   1-high='Child';

   value EARNINGSf 0='No Income'
                    0.00001-<450='Low Income'
                   450-<1200='Medium Income'
                   1200-high='High Income';

   value SMOKEf 0='No Smoke'
                0.001-high= 'Smoke';

   value SLEEPf low-<450 = 'Low Sleep'
                450-<600 = 'Medium Sleep'
                600-high = 'High Sleep';
```

```
   value EATf low-<30 = 'Low Eat'
                30-<90  = 'Medium Eat'
             90-high = 'High Eat';

   value TVf low-<30  = 'Low TV'
             30-<245  = 'Medium TV'
             245-high = 'High TV';

   value SPORTSf 0='No'
                  1='Yes';

run;
*;

********************************************************************
*;

* Create the starting dataset by reading the raw data file;

data start;

    * Note: length of the records in raw file is greater than
    * 256 bytes (default). In this case, record length is 3309,
    * so use LRECL=3309 to tell SAS to read a line up to 3309
characters;
    infile "C:\STAT 448\30901-0007-Data.txt" LRECL=3309;
    input

        /*Raw data file contains 412 variables; only selected
variables
           listed below were read. Also variable names different from
           the ones listed in the codebook*/

        /*Identifiers*/
        ID $1-14          /*TUCASEID-ATUS - Case ID*/
        DIARYDAY 174-181   /*TUDIARYDAY - Day of the week of diary day
                             (day of the week about which the
respondent
                              was interviewed)*/
        HOLIDAY 182-189    /*TRHOLIDAY - Flag to indicate if diary
                             day was a holiday*/

        /*Demographics*/
        AGE 38-45         /*TEAGE - Age*/
        SEX 46-53         /*TESEX - Sex*/
        RACE 62-69        /*PTDTRACE - Race*/
        HISPANIC 70-77    /*PEHSPNON - Hispanic or Latino?*/
        CHILD 150-157     /*TRCHILDNUM - Number of children < 18 lving
in the household*/
        STATUS 86-93      /*TELFS - Labor force status*/
        EARNINGS 142-149 /*TRERNWA - Weekly Earnings is defined for
all employed persons
```

```
                              who are not self-employed or without pay*/
         DEGREE 54-61       /*PEEDUCA - Education Attainment*/
         METRO 78-85        /*GTMETSTA - Metropolitan Status*/

         /*Activity: Sleeping*/
         T010101 190-197            T010102 198-205

            /*Activity: Eating and Drinking*/
         T110101 1678-1685        T110201 1686-1693

         /*Activity: Socializing and Communicating*/
         T120101 1694-1701

         /*Activity: Tobacco and drug use*/
         T120302 1734-1741

         /*Activity: Watching TV*/
          T120303 1742-1749        T120304 1750-1757

         /*Activity: Participating in Sports*/

          T130101 1926-1933        T130102 1934-1941        T130103
1942-1949
         T130104 1950-1957        T130105 1958-1965        T130106 1966-
1973
         T130107 1974-1981        T130109 1982-1989        T130110 1990-
1997
         T130112 1998-2005        T130113 2006-2013        T130114 2014-
2021
         T130116 2022-2029        T130117 2030-2037        T130118 2038-
2045
         T130119 2046-2053        T130120 2054-2061        T130122 2062-
2069
         T130123 2070-2077        T130124 2078-2085        T130125 2086-
2093
         T130126 2094-2101        T130127 2102-2109        T130128 2110-
2117
         T130129 2118-2125        T130130 2126-2133        T130131 2134-
2141
         T130132 2142-2149        T130133 2150-2157        T130134 2158-
2165
         T130136 2166-2173        T130199 2174-2181        T130301 2358-
2365;

run;
*;

************************************************************************
*;

* Create the main dataset by combining the activity variables,
* applying  formats and labels, and dealing with missing values;
```

```sas
data main;
    set start;
        /*Day of week indicator*/
        * If it is a weekend or holiday, weekday is 0;
        if DIARYDAY in (1,7) |  HOLIDAY = 1 then  WEEKDAY=0;
        else WEEKDAY=1;

        /*Convert earnings to appropriate units (implied 2 decimals)*/
        if earnings = -1 then earnings =0;
        EARNINGS=EARNINGS/100;

        /*Combine Activity variables*/
        TSLEEP = sum(of T01:);
        TEAT = sum(of T11:);
        TSOCIAL = T120101;
        TSMOKE = T120302;
        TTV = T120303 + T120304;
        TSPORTS = sum(of T1301:) + T130301;

    /*Drop unnecessary variables*/
    drop  ID DIARYDAY HOLIDAY T010101--T130301;

    *Apply labels;
    label
        ID         = 'ATUS Case ID'
        WEEKDAY   = 'Weekday (non-holiday)'
        AGE       = 'Age'
        SEX       = 'Sex'
        RACE      = 'Race'
        HISPANIC  = 'Hispanic'
        CHILD     = 'Number of children living in household'
        STATUS    = 'Labor force status'
        EARNINGS  = 'Weekly Earnings'
        DEGREE    = 'Education Attainment'
        METRO     = 'Metropolitan Status'
        TSLEEP    = 'Time Sleeping'
        TEAT      = 'Time Eating & Drinking'
        TSOCIAL   = 'Time Socializing & Communicating'
        TSMOKE    = 'Time Tobacco & Drug Use'
        TTV       = 'Time Watching TV'
        TSPORTS   = 'Time Participating in Sports';

    *Apply formats;
    format
        WEEKDAY  WEEKDAYf.
        SEX      SEXf.
        RACE      RACEf.
        HISPANIC HISPANICf.
        STATUS      STATUSf.
        DEGREE      DEGREEf.
        METRO     METROf.;
run;
*;
```

```sas
        ******************************************************************
        *;

        *Examine the descriptor portion of the SAS data set;
        proc contents data=main varnum;
        run;

        *Examine the data portion of the SAS data set;
        proc print data=main (firstobs=1 obs=10) noobs label;
            var AGE--TSPORTS;
        run;
        *;
        ******************************************************************
        *;

        *Subset data for weekdays only;

        data ATUSweek;
            set main;
            if WEEKDAY=1;
        run;
        *;

        ******************************************************************
        *;

        *Convert Time Activities to Categorical;

        data ATUScon;
            set ATUSweek;
            SLEEP = TSLEEP;
            EAT   = TEAT;
            TV    = TTV;
            SMOKE = TSMOKE;

            if TSPORTS > 0 then  SPORTS=1;
                else SPORTS=0;

            format
                SPORTS    SPORTSf.
                SLEEP     SLEEPf.
                EAT       EATf.
                TV        TVf.
                SMOKE      SMOKEf.;
            label
                SPORTS    = 'Participating in Sports'
                SLEEP     = 'Sleeping'
                EAT       = 'Eating & Drinking'
                SMOKE     = 'Tobacco & Drug Use'
                TV        = 'Watching TV';
        run;
        *;
```

```sas
*Convert Demographics to Categorical;

data ATUScat;
    set ATUScon;
    label
        CHILD = 'Children living in household';
    format
        AGE          AGEf.
        CHILD     CHILDf.
        EARNINGS EARNINGSf.;

run;
*;

*************************************************************************
*;

*Examine the descriptor portion of the SAS data set;
proc contents data=ATUScat varnum;
run;

*Examine the data portion of the SAS data set;
proc print data=ATUScat (firstobs=1 obs=10) noobs label;
    var AGE--SPORTS;
run;
*;

*************************************************************************
*;
*************************************************************************
*;
* Descriptive Analysis;

*3.1.1    Time Activities;
proc means data=ATUScat n mean median min max std maxdec=2;
    var TSPORTS TSLEEP TEAT TSOCIAL TTV TSMOKE;
run;
*;



*3.1.2    Time Spent on Sports;

* Participation in Sports;
proc sgplot data=atuscat;
    vbar SPORTS/ group=sports datalabel;
    xaxis label='Participated in Sports?';
run;
*;

* Statistics of TSPORTS;
proc univariate data=ATUScat normal;
    where TSPORTS>0;
```

```
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    probplot TSPORTS;
    ods select TestsForNormality BasicMeasures Moments
              Histogram ProbPlot;
run;

* 3.1.3    Time Spent on Sports by categorical variables;

* Histogram of TSPORTS by categorical variable;
proc univariate data=ATUScat;
    where TSPORTS>0;
    class SEX;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class AGE;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class RACE;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class HISPANIC;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class CHILD;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class STATUS;
```

```sas
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class EARNINGS;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class DEGREE;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class METRO;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class SMOKE;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class SLEEP;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;

proc univariate data=ATUScat;
    where TSPORTS>0;
    class TV;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;
```

```
proc univariate data=ATUScat;
    where TSPORTS>0;
    class EAT;
    var TSPORTS;
    histogram TSPORTS/ normal kernel;
    ods select Histogram;
run;
*;

* Times by sex;
proc sgplot data=ATUScat;
   vbar SEX/response=TSPORTS stat=mean datalabel
        fillattrs=(color=red)
                 transparency=.7;

   vbar SEX/response=TTV stat=mean datalabel
        fillattrs=(color=blue)
        transparency=.7 barwidth=.5;

   vbar SEX/response=TSLEEP stat=mean datalabel
            fillattrs=(color=green)
            transparency=.7 barwidth=.4;

   vbar SEX/response=TEAT stat=mean datalabel
            fillattrs=(color=orange)
            transparency=.7 barwidth=.6;
   yaxis label='Mean Time (min)';
run;

* Labor Satus by Degree;
proc sgpanel data=ATUScat;
    panelby DEGREE;
    where TSPORTS>0;
    vbox TSPORTS / category=STATUS group=STATUS;
    rowaxis label='Time Spent on Sports(min)';
run;

proc tabulate data=ATUScat;
    where TSPORTS>0;
    class SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
          METRO SMOKE;
    var TSPORTS;
    table DEGREE*STATUS, TSPORTS*(mean std n);
run;

* Race by Metro;
proc sgplot data=ATUScat;
    where TSPORTS>0;
    vbox TSPORTS / category=METRO group=RACE groupdisplay=cluster;
    yaxis label='Time Spent on Sports(min)';
run;

proc tabulate data=ATUScat;
```

```
    where TSPORTS>0;
    class SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
         METRO SMOKE;
    var TSPORTS;
    table METRO*RACE, TSPORTS*(mean std n);
run;

* Age by Sex;
proc sgpanel data=ATUScat;
    where TSPORTS>0;
    panelby SEX/layout=rowlattice novarname;
    hbox TSPORTS/category=AGE group=AGE;
    colaxis label='Time Spent on Sports(min)' grid;
run;

proc tabulate data=ATUScat;
    where TSPORTS>0;
    class SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
         METRO SMOKE;
    var TSPORTS;
    table SEX*AGE, TSPORTS*(mean std n);
run;

* Hispanic by Earnings;
proc sgplot data=ATUScat;
    where TSPORTS>0;
    vbox TSPORTS / category=EARNINGS group=HISPANIC
groupdisplay=cluster;
    yaxis label='Time Spent on Sports(min)';
run;

proc tabulate data=ATUScat;
    where TSPORTS>0;
    class SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
         METRO SMOKE;
    var TSPORTS;
    table EARNINGS*HISPANIC, TSPORTS*(mean std n);
run;

* Child by Smoke;
proc sgpanel data=ATUScat;
    where TSPORTS>0;
    panelby SMOKE/layout=rowlattice novarname;
    hbox TSPORTS/category=CHILD group=CHILD;
    colaxis label='Time Spent on Sports(min)' grid;
run;

proc tabulate data=ATUScat;
    where TSPORTS>0;
    class SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
         METRO SMOKE;
    var TSPORTS;
    table SMOKE*CHILD, TSPORTS*(mean std n);
```

```sas
run;
*;

*3.1.4    Participation in Sports by categorical variables;

* Frequencies of categorical variables;

proc freq data=ATUScat;
    table SEX AGE RACE HISPANIC CHILD STATUS EARNINGS DEGREE
          METRO SMOKE SLEEP EAT TV/nocum;
run;

* Age and Sex;
proc sgplot data=ATUScat;
  vbar AGE / response=SPORTS stat=sum group=SEX nostatlabel
         datalabel dataskin=sheen;
  xaxis display=(nolabel);
  yaxis grid;
run;

* Degree and Child;
proc sgpanel data=ATUScat;
    panelby DEGREE/spacing=5;
    vbar CHILD/stat=sum response=SPORTS group=CHILD nostatlabel
datalabel
        dataskin=sheen;
run;

* Earnings and Metro;
proc sgplot data=ATUScat;
  vbar EARNINGS / response=SPORTS stat=sum group=METRO nostatlabel
         datalabel dataskin=sheen groupdisplay=cluster;
  xaxis display=(nolabel);
  yaxis grid;
run;

* Race and smoke;
proc sgplot data=ATUScat;
  vbar SMOKE/ response=SPORTS stat=sum group=RACE nostatlabel
         datalabel dataskin=sheen;
  xaxis display=(nolabel);
  yaxis grid;
run;

* Status and Hispanic;
proc sgpanel data=ATUScat;
    panelby HISPANIC/spacing=5;
    vbar STATUS/stat=sum response=SPORTS group=STATUS nostatlabel
datalabel
        dataskin=sheen;
run;


*****************************************************************
```

```sas
*;
*******************************************************************
*;
* Associative Analysis;

* Scatter Plot;
proc sgscatter data=ATUScat;
    matrix TSLEEP--TSPORTS;
run;

* Test correlations;
proc corr data=ATUScat spearman;
    var TSLEEP TEAT TSOCIAL TTV TSPORTS;
    ods select SpearmanCorr ;
run;

proc corr data=ATUScat spearman;
    var SEX CHILD HISPANIC METRO SMOKE SPORTS;
    ods select SpearmanCorr ;
run;
*;

* Test differences in proportions among groups;
proc freq data=ATUScat;
    tables SPORTS*SEX / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*AGE / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*RACE / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*HISPANIC / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*CHILD / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*STATUS / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;
```

```sas
proc freq data=ATUScat;
    tables SPORTS*EARNINGS / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*DEGREE / chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*METRO/ chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*SMOKE/ chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*SLEEP/ chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*TV/ chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

proc freq data=ATUScat;
    tables SPORTS*EAT/ chisq nopercent norow;
    ods select ChiSq CrossTabFreqs;
run;

*;

*********************************************************************
*;
*********************************************************************
*;

* Inferential Analysis;
* Test differences in medians among groups;

proc npar1way data=ATUScat wilcoxon plots=all median;
    where TSPORTS>0;
    class SEX;
    var TSPORTS ;
    ods select KruskalWallisTest  WilcoxonBoxPlot ;
run;

proc npar1way data=ATUScat wilcoxon;
```

```sas
    where TSPORTS>0;
    class AGE;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class RACE;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class HISPANIC;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class CHILD;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class STATUS;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class EARNINGS;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class DEGREE;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class METRO;
    var TSPORTS;
    ods select KruskalWallisTest ;
```

```sas
run;


proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class SMOKE;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class SLEEP;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class TV;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;

proc npar1way data=ATUScat wilcoxon;
    where TSPORTS>0;
    class EAT;
    var TSPORTS;
    ods select KruskalWallisTest ;
run;
*;

**********************************************************************
*;
**********************************************************************
*;

*Predictive Analysis;

*Logistic Regression with Stepwise;

proc logistic data=ATUScat desc plots(MAXPOINTS=NONE)= all ;
  class SEX (ref='Female') AGE (ref='25 to 34 years ')
        RACE HISPANIC CHILD STATUS (ref='Employed')
        EARNINGS (ref='Low Income') DEGREE METRO
        SMOKE (ref='No Smoke') /param=ref;
  model SPORTS = SEX  AGE RACE HISPANIC CHILD STATUS
        EARNINGS DEGREE METRO SMOKE TSOCIAL TTV TSLEEP
TEAT/selection=stepwise lackfit ;

  oddsratio SEX/diff=ref;
  oddsratio AGE/diff=ref;
  oddsratio CHILD/diff=ref;
```

```
   oddsratio STATUS/diff=ref;
   oddsratio DEGREE/diff=ref;

   ods exclude ROCCurve effectplot;

run;


* Logistic Regression with selected model;
* Need to do separate to rerun proc logistic to:
   * Get desired effectplots;
   * Get all pairs of odds ratio;

proc logistic data=ATUScat desc;
   class SEX (ref='Female') AGE (ref='25 to 34 years ')
         RACE HISPANIC CHILD STATUS (ref='Employed')
         EARNINGS (ref='Low Income') DEGREE METRO
         SMOKE (ref='No Smoke') /param=ref;
   model SPORTS = SEX AGE  CHILD STATUS
          DEGREE TTV TSLEEP;

   oddsratio SEX;
   oddsratio AGE;
   oddsratio CHILD;
   oddsratio STATUS;
   oddsratio DEGREE;

   effectplot slicefit(x=TTV sliceby= age) / at(DEGREE = 'Bachelors
degree' SEX='Male') noobs;
   effectplot slicefit(x=TSLEEP sliceby= status) / at(DEGREE =
'Bachelors degree' SEX='Male')  noobs;
   effectplot slicefit(x=TSLEEP sliceby= degree) / at( SEX='Male')
noobs;
   effectplot fit(plotby=SEX X=TTV)/at(DEGREE = 'Bachelors degree')
noobs;
   effectplot fit(plotby=CHILD X=TSLEEP)/at(DEGREE = 'Bachelors
degree' SEX='Male') noobs;

run;

* rtf;
ods rtf close;
```