

Topics currently being or that can be analyzed with WSPP – 3

1. What is the impact of smoking by “best friends” on youth smoking?
2. What are the effects of school level smoking on youth smoking?
3. What are the effects of parental and sibling smoking habits?
4. What is the effect of early initiation on youth smoking?
5. What are the effects of higher taxes on youth smoking?
6. Does youth smoking vary according to geography?

Topics that cannot be analyzed due to lack of data

1. Does youth smoking vary according to socioeconomic status?
2. Is educational attainment correlated with smoking?
3. Is health correlated with youth smoking?

Code for the WSPP3 Data set

```
options nocenter errors=2;
libname data ".";
```

```
*now reading in data from survey;
```

```
filename in1 "nevil.sas";
data dat1;
infile in1;
input id gender grade age bmonth bday byear q4 q5 q6 q7 q8 q9
      q10 q11 q12 q13 q14 q15 q16 q17 q18 q19 q20 q21 q22 q23 q24 q25
      q33 q34 q35 q36 q38 q39 q40 q41 q42 q43 q44 q45 q46 q47 q48 smoke;
run;
```

```
proc contents;
run;
```

```
data one;
set dat1;
keep id gender grade age q7 q8 q9 q10 q16 q19 q20 q21 q22 q23 q24 q33
byear q25 q16 q13 q34 q36 q35 q38 q39 q40 q41 q44 q46 q47;
run;
```

```
*cleaning up the data- basically deleting missing observations;
```

```
data one;
set one;
if age=. then delete;
if age=88 or age=99 then delete;
if gender=. then delete;
if q7=. or q7=88 or q7=99 then delete;
if q8=. or q8=8888 or q8=9999 then delete;
if q9=. or q9=8888 or q9=9999 then delete;
if q10=. or q10=8888 or q10=9999 then delete;
if q23=. then q23=0;
if byear=. then delete;
if byear=99 then delete;
if q33=. or q33=8 or q33=9 then q33=0;
if q25=. or q25=88 or q25=99 then q25=0;
if q33=5 or q33=4 then univ=1;else univ=0;
if q33=3 or q33=2 then tech=1;else tech=0;
if q33=1 then job=1;else job=0;
hmdaily=q8;
hmweekly=q9;
hmmmonthly=q10;
```

dad=q19;
mom=q20;
bro=q21;
sis=q22;
frnd=q23;
bfrnd=q24;
*please notice this capping daily consumption;
if hmdaily gt 100 then hmdaily=100;
if grade=6 then year=1990;
if grade=7 then year=1991;
if grade=8 then year=1992;
if grade=9 then year=1993;
if grade=10 then year=1994;
if grade=11 then year=1995;
if grade=12 then year=1996;
if grade=6 then year1=90;
if grade=7 then year1=91;
if grade=8 then year1=92;
if grade=9 then year1=93;
if grade=10 then year1=94;
if grade=11 then year1=95;
if grade=12 then year1=96;
*cigprice is average ontario cigarette price;

if year=1990 then cigprice=31.6079;
if year=1991 then cigprice=41.6417;
if year=1992 then cigprice=46.0829;
if year=1993 then cigprice=45.8833;
if year=1994 then cigprice=25.7126;
if year=1995 then cigprice=24.6067;
if year=1996 then cigprice=25.0214;

if year=1990 then cig1=27.0021;
if year=1991 then cig1=31.6079;
if year=1992 then cig1=41.6417;
if year=1993 then cig1=46.0829;
if year=1994 then cig1=45.8833;
if year=1995 then cig1=25.7126;
if year=1996 then cig1=24.6067;

if year=1990 then cig2=24.0077;
if year=1991 then cig2=27.0021;
if year=1992 then cig2=31.6079;
if year=1993 then cig2=41.6417;
if year=1994 then cig2=46.0829;
if year=1995 then cig2=45.8833;

if year=1996 then cig2=25.7126;

if year=1990 then cigtax=21.3922;
if year=1991 then cigtax=30.7474;
if year=1992 then cigtax=33.1777;
if year=1993 then cigtax=33.1777;
if year=1994 then cigtax=12.9375;
if year=1995 then cigtax=11.7300;
if year=1996 then cigtax=12.0558;

if year=1990 then cigt1=18.4923;
if year=1991 then cigt1=21.3922;
if year=1992 then cigt1=30.7474;
if year=1993 then cigt1=33.1777;
if year=1994 then cigt1=33.1777;
if year=1995 then cigt1=12.9375;
if year=1996 then cigt1=11.7300;

if year=1990 then cigt2=15.5246;
if year=1991 then cigt2=18.4923;
if year=1992 then cigt2=21.3922;
if year=1993 then cigt2=30.7474;
if year=1994 then cigt2=33.1777;
if year=1995 then cigt2=33.1777;
if year=1996 then cigt2=12.9375;

if year=1990 then alcprice=106.2;
if year=1991 then alcprice=108.9;
if year=1992 then alcprice=100.0;
if year=1993 then alcprice=115.8;
if year=1994 then alcprice=117.5;
if year=1995 then alcprice=121.8;
if year=1996 then alcprice=123.5;
if year=1990 then unemp=6.2;
if year=1991 then unemp=9.5;
if year=1992 then unemp=10.7;
if year=1993 then unemp=10.9;
if year=1994 then unemp=9.6;
if year=1995 then unemp=8.7;
if year=1996 then unemp=8.1;
agesq=age*age;
lcig=log(cigprice);

```

lage=log(age);
lagesq=log(agesq);
if hmdaily=0 then hmdaily=0.00000001;
lhmdail=log(hmdaily);
lcig1=log(cig1);
lcig2=log(cig2);
lcigt1=log(cigt1);
lcigt2=log(cigt2);
if gender=. then gender=0;
if q16=4 or q16=5 then future=1;else future=0;
if q13=. then q13=8;
if q13=4 or q13=3 then quit=1;if q13=1 or q13=2 or q13=8 or q13=9 then
quit=2;if q13=5 then quit=0;
if q34=. then q34=0;if q36=. then q36=0;if q35=. then q35=0;
if q34=1 then parent=1;else parent=0;
if q34=2 then relative=1;else relative=0;
if q34=3 then friendp=1;else friendp=0;
if q34=4 then own=1;else own=0;
if q36 le 3 and q36 ge 1 then work=1;else work=0;
if q36=1 then tban=1;else tban=0;
if q36=2 then banin=1;else banin=0;
if q36=3 or q36=4 then noban=1;else noban=0;
if q35 ge 1 and q35 le 3 then apart=1;else apart=0;
if q35=1 then tbanap=1; else tbanap=0;
if q35=2 then baninap=1; else baninap=0;
if q35=3 then someap=1; else someap=0;
if q35=4 then norestrict=1; else norestrict=0;
if q41=. then q41=0;if q40=. then q40=0;if q44=. then q44=0;
if q46=. then q46=0;if q47=. then q47=0;
if q38=. then q38=0;if q39=. then q39=0;
if q38=1 or q38=2 then double=1;else double=0;
if q39=1 or q39=2 then add=1;else add=0;
if q40=1 or q40=2 then group=1;else group=0;
if q41=1 or q41=2 then audio=1;else audio=0;
if q44=1 or q44=2 then assembly=1;else assembly=0;
if q46=1 or q46=2 then pplace=1;else pplace=0;
if q47=1 or q47=2 then illegal=1;else illegal=0;
help=1;
run;

*proc print;
*var id grade scode;
*run;

```

*q7 is a categorical variable denoting the level of smoking. In terms of

daily smoking q8. We want to creat a variable If you smoke or not;

```
data one;
set one;
if q7 gt 3 then apdys=1;else apdys=0;
if q7 gt 8 then daily=1;else daily=0;
run;
```

```
proc sort data=one;
by id year;
```

*for some reason many duplicate observations;

```
proc sort nodupkeys;
by id year apdys;
run;
```

*now to create variable denoting number of smokers in each class;

```
data school;
set one;
yo=1;
run;
```

```
proc print;
var id scode grade apdys;
run;
```

```
proc sort;
by scode grade;
```

```
proc means;
```

```
proc means;
var yo apdys;
by scode grade;
output out=scode sum=;
run;
```

```
data scode;
set scode;
if scode=. or grade=. then delete;
if apdys=. or yo=. then delete;
agcigsc=apdys;
```

```
yotot=yo;  
keep agcigsc scode grade yotot;  
run;  
proc print;  
run;
```

*yotot is total number of classmates, agcigsc is number of classmates who smoke;

```
proc sort data=one;  
by scode grade;  
run;
```

```
proc sort data=scode;  
by scode grade;  
run;
```

```
data one;  
merge one scode;  
by scode grade;  
pclsmok= agcigsc/yotot;  
run;
```

```
proc means;  
var pclsmok agcigsc yotot;  
run;
```

*pclsmok is proportion of classmates who smoke;

```
proc sort;  
by id year;  
run;
```

*how may cigs per day for daily smokers;

```
data two;  
set one;  
if daily=0 then delete;  
run;
```

```
proc means;  
var hmdaily;  
run;
```

```
data three;
```

```

set one;
*these will be used as family fixed effects;
if q19=4 then dad=1; else dad=0;
if q20=4 then mom=1; else mom=0;
if q21=1 then bro=1; else bro=0;
if q22=1 then sis=1; else sis=0;
frnd=q23;
if frnd gt 6 then frnd=0;
if frnd=0 then frnd=0.00000001;
lfrnd=log(frnd);
if q24=2 then bfrnd=1; else bfrnd=0;
bfno=frnd*bfrnd;
if dad=1 or mom=1 then one=1;else one=0;
if dad=1 and mom=1 then both=1;else both=0;
if bro=1 or sis=1 then sib=1;else sib=0;
*create grade fixed effects;
if grade=6 then g6=1; else g6=0;
if grade=7 then g7=1; else g7=0;
if grade=8 then g8=1; else g8=0;
if grade=9 then g9=1; else g9=0;
if grade=10 then g10=1; else g10=0;
if grade=11 then g11=1; else g11=0;
if grade=12 then g12=1; else g12=0;
run;

```

```

proc means;
var dad mom bro sis frnd bfrnd;
run;

```

```

*create data for analysis from grade 8-12;
data four;
set one;
lagfr=lag(frnd);
lagfrr=lag(lagfr);
bfno=frnd*bfrnd;
if grade=6 then delete;
if grade=7 then delete;
if grade=8 then g8=1; else g8=0;
if grade=9 then g9=1; else g9=0;
if grade=10 then g10=1; else g10=0;
if grade=11 then g11=1; else g11=0;
if grade=12 then g12=1; else g12=0;
*these will be used as family fixed effects;
if q19=4 then dad=1; else dad=0;
if q20=4 then mom=1; else mom=0;
if q21=1 then bro=1; else bro=0;

```

```
if q22=1 then sis=1; else sis=0;
frnd=q23;
if frnd gt 6 then frnd=0;
if frnd=0 then frnd=0.00000001;
lfrnd=log(frnd);
if q24=2 then bfrnd=1; else bfrnd=0;
if dad=1 or mom=1 then one=1;else one=0;
if dad=1 and mom=1 then both=1;else both=0;
```

```
if bro=1 or sis=1 then sib=1;else sib=0;
run;
```

```
proc print;
var id frnd lagfr;
run;
```