
log: c:\NHANES\log\linear.log
log type: text
opened on: 4 Aug 2008, 11:25:06

. use "C:\NHANES\DATA\analysis_data.dta", clear

. ***create variable codes*****
. capture label drop educ

. gen educ= dmdeduc
(3462 missing values generated)

. replace educ=. if educ >3
(44 real changes made, 44 to missing)

. gen smoker=1 if smq020==2
(15637 missing values generated)

. replace smoker=2 if smq020==1 & smq040==3
(2724 real changes made)

. replace smoker=3 if smq020==1 & (smq040==1 | smq040==2)
(2170 real changes made)

. gen bmicat=1 if bmx bmi >=0 & bmx bmi <18.5
(17446 missing values generated)

. replace bmicat=2 if bmx bmi >=18.5 & bmx bmi <25
(6084 real changes made)

. replace bmicat=3 if bmx bmi >=25 & bmx bmi <30
(4295 real changes made)

. replace bmicat=4 if bmx bmi >=30 & bmx bmi <.
(3535 real changes made)

. /*eligible is 1 if other variables used in final regression model are non-bla
> nk; this keeps #obs same*/

. gen eligible=1 if (ridageyr >=20 & ridageyr <.) & (bmx bmi !=. & riagendr !=.
> & ridreth1 !=. & ridageyr !=. & smoker !=. & educ !=.)
(11967 missing values generated)

. tab bmicat if eligible==1, gen(Ibmicat)

bmicat	Freq.	Percent	Cum.
-----+-----			
1	151	1.67	1.67 ₁

2	2,758	30.52	32.19
3	3,279	36.28	68.47
4	2,849	31.53	100.00

Total	9,037	100.00	
-------	-------	--------	--

```

.
. ****format variables*****
. capture label drop sexfmt

. label define sexfmt 1 "male"

. label define sexfmt 2 "female", add

.
. label define race2fmt 1 "Mex American"

. label define race2fmt 3 "NH White", add

. label define race2fmt 4 "NH Black", add

. capture label drop smkfmt

. label define smkfmt 1 "Never smoker"

. label define smkfmt 2 "Past smoker", add

. label define smkfmt 3 "Current smoker", add

.
. label define educ 1 "< High school"

. label define educ 2 "High school", add

. label define educ 3 "> High school", add

.
. label define bmicat 1 "underweight"

. label define bmicat 2 "normal weight", add

. label define bmicat 3 "overweight", add

. label define bmicat 4 "obese", add

.
. label values riagendr sexfmt

. label values ridreth1 race2fmt

. label values smoker smkfmt

. label values educ educ

```

```
. label values bmicat bmicat
.
. ****specify survey design variables****
. svyset sdmvpsu [pweight=wtmec4yr], strata(sdmvstra) vce(linearized)
```

```
    pweight: wtmec4yr
      VCE: linearized
Single unit: missing
  Strata 1: sdmvstra
    SU 1: sdmvpsu
    FPC 1: <zero>
```

```
. ****simple linear regression***
. svy, subpop(if eligible==1): regress lbdhdl bmxbmi
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata	=	28	Number of obs	=	20462
Number of PSUs	=	57	Population size	=	268632486
			Subpop. no. of obs	=	8495
			Subpop. size	=	181112209
			Design df	=	29
			F(1, 29)	=	411.06
			Prob > F	=	0.0000
			R-squared	=	0.0773

		Linearized				
lbdhdl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
bmxbmi	-.688685	.0339678	-20.27	0.000	-.7581569	-.6192131
_cons	70.54016	1.095367	64.40	0.000	68.29988	72.78044

```
. svy, subpop(if eligible==1): regress lbdhdl bmicat
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata	=	28	Number of obs	=	20462
Number of PSUs	=	57	Population size	=	268632486
			Subpop. no. of obs	=	8495
			Subpop. size	=	181112209
			Design df	=	29
			F(1, 29)	=	670.46
			Prob > F	=	0.0000
			R-squared	=	0.0901

		Linearized				
lbdhdl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	

```

-----+-----
      bmicat |   -5.554894   .2145314  -25.89   0.000   -5.99366   -5.116128
      _cons |   67.52966   .8129439   83.07   0.000    65.867    69.19231
-----+-----

```

```

. svy, subpop(if eligible==1): mean lbdhdl, over(bmicat)
(running mean on estimation sample)

```

Survey: Mean estimation

```

Number of strata =      28      Number of obs   =      15332
Number of PSUs   =      57      Population size = 237013191
                                   Subpop. no. obs  =      8495
                                   Subpop. size    = 181112209
                                   Design df       =      29

```

```

underweight: bmicat = underweight
  _subpop_2: bmicat = normal weight
overweight: bmicat = overweight
  obese: bmicat = obese

```

```

-----+-----
      Over |           Linearized
           |           Mean   Std. Err.   [95% Conf. Interval]
-----+-----
lbdhdl   |
underweight |   60.43627   1.497248   57.37405   63.49849
  _subpop_2 |   57.17528   .5253125   56.1009   58.24966
overweight |   49.69381   .3696679   48.93775   50.44986
  obese    |   45.93693   .4375764   45.04198   46.83187
-----+-----

```

```

. svy, subpop(if eligible==1): mean lbdhdl, over(riagendr)
(running mean on estimation sample)

```

Survey: Mean estimation

```

Number of strata =      28      Number of obs   =      15711
Number of PSUs   =      57      Population size = 242760411
                                   Subpop. no. obs  =      8495
                                   Subpop. size    = 181112209
                                   Design df       =      29

```

```

male: riagendr = male
female: riagendr = female

```

```

-----+-----
      Over |           Linearized
           |           Mean   Std. Err.   [95% Conf. Interval]
-----+-----
lbdhdl   |
male    |   45.91329   .3367541   45.22455   46.60203
female  |   56.21287   .4938749   55.20279   57.22296
-----+-----

```

```
. svy, subpop(if eligible==1): mean lbdhdl, over(ridreth1)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      28      Number of obs   =      15711
Number of PSUs   =      57      Population size = 242760411
                                   Subpop. no. obs  =      8495
                                   Subpop. size     = 181112209
                                   Design df        =      29
```

```
_subpop_1: ridreth1 = Mex American
      2: ridreth1 = 2
_subpop_3: ridreth1 = NH White
_subpop_4: ridreth1 = NH Black
      5: ridreth1 = 5
```

```
-----+-----
```

	Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
lbdhdl					
_subpop_1		48.92301	.2949188	48.31984	49.52619
2		47.70808	.6661662	46.34562	49.07055
_subpop_3		51.38278	.5050055	50.34993	52.41563
_subpop_4		54.50162	.5140192	53.45033	55.5529
5		50.91074	1.407939	48.03118	53.7903

```
-----+-----
```

```
. svy, subpop(if eligible==1): mean lbdhdl, over(smoker)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      28      Number of obs   =      8792
Number of PSUs   =      57      Population size = 186252458
                                   Subpop. no. obs  =      8495
                                   Subpop. size     = 181112209
                                   Design df        =      29
```

```
_subpop_1: smoker = Never smoker
_subpop_2: smoker = Past smoker
_subpop_3: smoker = Current smoker
```

```
-----+-----
```

	Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
lbdhdl					
_subpop_1		52.03534	.4235226	51.16914	52.90154
_subpop_2		51.59248	.5627116	50.44161	52.74336
_subpop_3		49.2751	.6252691	47.99628	50.55392

```
-----+-----
```

```
. svy, subpop(if eligible==1): mean lbdhdl, over(educ)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      28      Number of obs      =      14939
Number of PSUs   =      57      Population size   = 235230189
                                   Subpop. no. obs    =      8495
                                   Subpop. size      = 181112209
                                   Design df         =      29
```

```
_subpop_1: educ = < High school
_subpop_2: educ = High school
_subpop_3: educ = > High school
```

```
-----+-----
```

	Over	Mean	Std. Err.	[95% Conf. Interval]	
lbdhdl					
_subpop_1		49.37097	.4672992	48.41524	50.32671
_subpop_2		50.29367	.5427691	49.18358	51.40376
_subpop_3		52.47131	.4486635	51.55369	53.38893

```
-----+-----
```

```
.
. ****change reference level for categorical variables in model*****
. char ridreth1[omit]3
```

```
. char smoker[omit]3
```

```
. char educ[omit]3
```

```
. char bmicat[omit]2
```

```
.
.
. ****NOTE: xi and i.var are used to expand and denote categorical variables***
> **
```

```
. xi:svy, subpop(if eligible==1): regress lbdhdl i.bmicat
i.bmicat      _Ibmicat_1-4      (naturally coded; _Ibmicat_2 omitted)
(running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata =      28      Number of obs      =      20462
Number of PSUs   =      57      Population size     = 268632486
                                   Subpop. no. of obs   =      8495
                                   Subpop. size          = 181112209
                                   Design df             =      29
                                   F( 3, 27)              =      211.84
                                   Prob > F              =      0.0000
                                   R2squared              =      0.0935
```

lbdhd1	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Ibmicat_1	3.26099	1.412836	2.31	0.028	.3714158	6.150564
_Ibmicat_3	-7.481474	.4970235	-15.05	0.000	-8.498002	-6.464947
_Ibmicat_4	-11.23835	.4531955	-24.80	0.000	-12.16524	-10.31147
_cons	57.17528	.5253125	108.84	0.000	56.1009	58.24966

. ****multiple linear regression***

. xi: svy, subpop(if ridageyr >=20 & ridageyr <.): regress lbdhd1 ///

> i.riagendr i.ridreth1 ridageyr i.smoker i.educ i.bmicat

i.riagendr _Iriagendr_1-2 (naturally coded; _Iriagendr_1 omitted)

i.ridreth1 _Iridreth1_1-5 (naturally coded; _Iridreth1_3 omitted)

i.smoker _Ismoker_1-3 (naturally coded; _Ismoker_3 omitted)

i.educ _Ieduc_1-3 (naturally coded; _Ieduc_3 omitted)

i.bmicat _Ibmicat_1-4 (naturally coded; _Ibmicat_2 omitted)

(running regress on estimation sample)

Survey: Linear regression

Number of strata	=	28	Number of obs	=	20028
Number of PSUs	=	57	Population size	=	261345193
			Subpop. no. of obs	=	8495
			Subpop. size	=	181112209
			Design df	=	29
			F(13, 17)	=	219.66
			Prob > F	=	0.0000
			R-squared	=	0.2324

lbdhd1	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iriagendr_2	9.97728	.3228816	30.90	0.000	9.316913	10.63765
_Iridreth1_1	.6690456	.5512029	1.21	0.235	-.4582909	1.796382
_Iridreth1_2	-2.243026	.6598387	-3.40	0.002	-3.592548	-.8935046
_Iridreth1_4	4.950362	.6551325	7.56	0.000	3.610465	6.290258
_Iridreth1_5	-.5570998	1.109527	-0.50	0.619	-2.826338	1.712138
ridageyr	.1127225	.0127682	8.83	0.000	.0866086	.1388364
_Ismoker_1	1.220217	.7059541	1.73	0.095	-.2236216	2.664055
_Ismoker_2	2.333292	.6502412	3.59	0.001	1.0034	3.663185
_Ieduc_1	-3.030584	.5535365	-5.47	0.000	-4.162693	-1.898475
_Ieduc_2	-1.854561	.5513662	-3.36	0.002	-2.982232	-.7268909
_Ibmicat_1	2.298258	1.339698	1.72	0.097	-.4417327	5.038249
_Ibmicat_3	-6.545418	.4088177	-16.01	0.000	-7.381545	-5.709292
_Ibmicat_4	-12.00343	.3834687	-31.30	0.000	-12.78771	-11.21915
_cons	46.3097	.5733786	80.77	0.000	45.13701	47.48239

```
. *** NOTE: The -adjust- command uses only the sample mean, not the mean based
> on the survey design, ***
. *** when performing its computations. Therefore, if you want to use the surv
> ey mean, ***
. *** you would need to calculate it first and specify it explicitly in the -ad
> just- command. ***
. *** The following commands use -summarize- which is an rclass command and wil
> l not cause any ***
. *** trouble if run between the -svy: regress- and -adjust- commands whereas -
> svy:mean- is an ***
. *** eclass command and cannot be used in between these commands.
> ***
```

```
. ****get survey means within the subpopulation****
. quietly {
```

```
. ****Use adjust command to produced adjusted means****
. adjust _Iridreth1_1=`rid1' _Iridreth1_2=`rid2' _Iridreth1_4=`rid4' ///
> _Iridreth1_5=`rid5' ridageyr=`ridage' _Ismoker_1=`smoke1' ///
> _Ismoker_2=`smoke2' _Ieduc_1=`educ1' _Ieduc_2=`educ2' ///
> _Ibmicat_1=`bmicat1' _Ibmicat_3=`bmicat3' _Ibmicat_4=`bmicat4' ///
> if ridageyr >=20 & ridageyr <. & e(sample), by(riagendr) se
```

```
-----
Dependent variable: lbdhdl      Command: regress
Variable left as is: _Iriagendr_2
Covariates set to value: _Iridreth1_1 = .07127677, _Iridreth1_2 = .06760141,
                        _Iridreth1_4 = .10102575, _Iridreth1_5 = .0424189,
                        ridageyr = 45.754149, _Ismoker_1 = .50536571,
                        _Ismoker_2 = .25203879, _Ieduc_1 = .21241539,
                        _Ieduc_2 = .25653796, _Ibmicat_1 = .01996474,
                        _Ibmicat_3 = .34825344, _Ibmicat_4 = .30082995
-----
```

```
-----
gender - |
adjudicat |
ed.      |          xb          stdp
-----+-----
      male |      46.0804      (.287663)
      female |      56.0577      (.410735)
-----
```

```
Key:  xb   = Linear Prediction
      stdp = Standard Error
```

```
. adjust _Iriagendr_2=`riagendr2' ridageyr=`ridage' _Ismoker_1=`smoke1' ///
> _Ismoker_2=`smoke2' _Ieduc_1=`educ1' _Ieduc_2=`educ2' ///
> _Ibmicat_1=`bmicat1' _Ibmicat_3=`bmicat3' _Ibmicat_4=`bmicat4' ///
> if ridageyr >=20 & ridageyr <. & e(sample), by(ridreth1) se
```


Dependent variable: lbdhdl Command: regress

Variables left as is: _Iridreth1_1, _Iridreth1_2, _Iridreth1_4, _Iridreth1_5

Covariates set to value: _Iriagendr_2 = .51854628, ridageyr = 45.754149,
_Ismoker_1 = .50536571, _Ismoker_2 = .25203879,
_Ieduc_1 = .21241539, _Ieduc_2 = .25653796,
_Ibmicat_1 = .01996474, _Ibmicat_3 = .34825344,
_Ibmicat_4 = .30082995

```
-----  
-----  
race/ethnicity - recode |          xb          stdp  
-----+-----  
Mex American |      51.5506   (.367634)  
      2 |      48.6385   (.525098)  
NH White |      50.8816   (.420078)  
NH Black |      55.8319   (.495734)  
      5 |      50.3245   (1.06718)
```

Key: xb = Linear Prediction
stdp = Standard Error

```
.  
. adjust _Iriagendr_2=`riagendr2' _Iridreth1_1=`rid1' _Iridreth1_2=`rid2' ///  
> _Iridreth1_4=`rid4' _Iridreth1_5=`rid5' ridageyr=`ridage' ///  
> _Ieduc_1=`educ1' _Ieduc_2=`educ2' _Ibmicat_1=`bmicat1' ///  
> _Ibmicat_3=`bmicat3' _Ibmicat_4=`bmicat4' if ridageyr >=20 & ///  
> ridageyr <. & e(sample), by(smoker) se
```

Dependent variable: lbdhdl Command: regress

Variables left as is: _Ismoker_1, _Ismoker_2

Covariates set to value: _Iriagendr_2 = .51854628, _Iridreth1_1 = .07127677,
_Iridreth1_2 = .06760141, _Iridreth1_4 = .10102575,
_Iridreth1_5 = .0424189, ridageyr = 45.754149,
_Ieduc_1 = .21241539, _Ieduc_2 = .25653796,
_Ibmicat_1 = .01996474, _Ibmicat_3 = .34825344,
_Ibmicat_4 = .30082995

```
-----  
-----  
smoker |          xb          stdp  
-----+-----  
Never smoker |      51.2696   (.370927)  
Past smoker |      52.3827   (.418068)  
Current smoker |      50.0494   (.630576)
```

Key: xb = Linear Prediction
stdp = Standard Error

```
.  
. adjust _Iriagendr_2=`riagendr2' _Iridreth1_1=`rid1' _Iridreth1_2=`rid2' ///  
> _Iridreth1_4=`rid4' _Iridreth1_5=`rid5' ridageyr=`ridage' ///  
> _Ismoker_1=`smoke1' _Ismoker_2=`smoke2' _Ibmicat_1=`bmicat1' ///
```

```
> _Ibmicat_3=`bmicat3' _Ibmicat_4=`bmicat4' if ridageyr >=20 & ///
> ridageyr <. & e(sample), by(educ) se
```

```
-----
Dependent variable: lbdhdl      Command: regress
Variables left as is: _Ieduc_1, _Ieduc_2
Covariates set to value: _Iriagendr_2 = .51854628, _Iridreth1_1 = .07127677,
                        _Iridreth1_2 = .06760141, _Iridreth1_4 = .10102575,
                        _Iridreth1_5 = .0424189, ridageyr = 45.754149,
                        _Ismoker_1 = .50536571, _Ismoker_2 = .25203879,
                        _Ibmicat_1 = .01996474, _Ibmicat_3 = .34825344,
                        _Ibmicat_4 = .30082995
-----
```

```
-----
```

	educ	xb	stdp
<	High school	49.343	(.452779)
	High school	50.519	(.513191)
>	High school	52.3736	(.377951)

```
-----
```

```
Key:  xb    = Linear Prediction
      stdp   = Standard Error
```

```
.
. adjust _Iriagendr_2=`riagendr2' _Iridreth1_1=`rid1' _Iridreth1_2=`rid2' ///
> _Iridreth1_4=`rid4' _Iridreth1_5=`rid5' ridageyr=`ridage' ///
> _Ieduc_1=`educ1' _Ieduc_2=`educ2' _Ismoker_1=`smoke1' ///
> _Ismoker_2=`smoke2' if ridageyr >=20 & ridageyr <. & e(sample), by(bmicat) s
> e
```

```
-----
Dependent variable: lbdhdl      Command: regress
Variables left as is: _Ibmicat_1, _Ibmicat_3, _Ibmicat_4
Covariates set to value: _Iriagendr_2 = .51854628, _Iridreth1_1 = .07127677,
                        _Iridreth1_2 = .06760141, _Iridreth1_4 = .10102575,
                        _Iridreth1_5 = .0424189, ridageyr = 45.754149,
                        _Ieduc_1 = .21241539, _Ieduc_2 = .25653796,
                        _Ismoker_1 = .50536571, _Ismoker_2 = .25203879
-----
```

```
-----
```

	bmicat	xb	stdp
	underweight	59.3969	(1.36473)
	normal weight	57.0987	(.401431)
	overweight	50.5533	(.375893)
	obese	45.0952	(.385962)

```
-----
```

```
Key:  xb    = Linear Prediction
      stdp   = Standard Error
```

```
.
. ***Model using an interaction term***
```

```

. **xi: svy, subpop(if ridageyr >=20 & ridageyr <.): regress lbdhdl i.riagendr
> i.ridreth1 ridageyr i.smoker i.educ i.smoker*bmxbmi**
.
. ****test all coeffiecients in model together****
. **test**
.
. ****test each coefficient in model separately***
. **test _Iriagendr_2, nosvyadjust**
. **test _Iridreth1_1 _Iridreth1_2 _Iridreth1_4 _Iridreth1_5, nosvyadjust**
. **test ridageyr, nosvyadjust**
. **test _Ibmicat_1 _Ibmicat_3 _Ibmicat_4, nosvyadjust**
. **test _Ismoker_1 _Ismoker_2, nosvyadjust**
. **test _Ieduc_1 _Ieduc_2, nosvyadjust**
.
. ****test never smokers vs. past smokers****
. **test _Ismoker_1 - _Ismoker_2 =0**
.
. ***Or, can use: test _Ismoker_1 = _Ismoker_2 ****
.
. ****
.
. **log close**
.
.
end of do-file

```