```
                                    ___  ____  ____  ____  ____ (R)
                                   /__    /   ____/   /    ____/
                                  ___/   /   /___/   /    /___/
                                    Statistics/Data Analysis
```
_____

```
        name:  <unnamed>
         log:  C:\Users\ramon\Desktop\UZH\Empirical Methods\Problem Sets\Problem Set 1\Stata\log_gm
    log type:  smcl
   opened on:  21 Oct 2019, 15:13:29
```

```
 1 .
 2 . use "C:\Users\ramon\Desktop\UZH\Empirical Methods\Problem Sets\Problem Set 1\Stata\smoke.dta"

 3 .
 4 . *8a) How many obs
 5 .
 6 . display _N
   807

 7 .
 8 . *8b) Summary statistics for cigs, educ, age, income, white, restaurn
 9 .
10 . asdoc sum cigs educ age income white restaurn
   (File Myfile.doc already exists, option append was assumed)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
cigs	807	8.686493	13.72152	0	80
educ	807	12.47088	3.057161	6	18
age	807	41.23792	17.02729	17	88
income	807	19304.83	9142.958	500	30000
white	807	.8785626	.3268375	0	1
restaurn	807	.2465923	.4312946	0	1

```
   Click to Open File:  Myfile.doc
```

```
11 .
12 . *8c)
13 . **i) Compute Beta 1 and 2 (error in the PS, we use 1 and 2, not 0 and 1)
14 . ***B2)
15 . gen COV = 0

16 . correlate educ cigs, covariance
   (obs=807)
```

	educ	cigs
educ	9.34624	
cigs	-2.04264	188.28

```
17 . replace COV = r(cov_12)
   (807 real changes made)

18 .
19 . egen SD = sd(educ)

20 . gen VAR = SD^2

21 .
22 . display COV/VAR
   -.21855212

23 .
24 . ***B1)
```

```
25 . gen B2 = COV/VAR

26 .
27 . gen mean_cigs = 0

28 . mean(cigs)

   Mean estimation                    Number of obs   =        807
```

	Mean	Std. Err.	[95% Conf. Interval]	
cigs	8.686493	.4830202	7.738367	9.634619

```
29 . matrix b=e(b)

30 . replace mean_cigs=b[1,1]
   (807 real changes made)

31 .
32 . gen mean_educ = 0

33 . mean(educ)

   Mean estimation                    Number of obs   =        807
```

	Mean	Std. Err.	[95% Conf. Interval]	
educ	12.47088	.1076172	12.25964	12.68212

```
34 . matrix b=e(b)

35 . replace mean_educ=b[1,1]
   (807 real changes made)

36 .
37 . gen B1 = mean_cigs - mean_educ*B2

38 .
39 . display B1
   11.41203

40 .
41 . **ii) Regression
42 .
43 . reg cigs educ
```

Source	SS	df	MS	Number of obs	=	807
				F(1, 805)	=	1.91
Model	359.817074	1	359.817074	Prob > F	=	0.1670
Residual	151393.866	805	188.066914	R-squared	=	0.0024
				Adj R-squared	=	0.0011
Total	151753.683	806	188.280003	Root MSE	=	13.714

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ	-.2185521	.1580048	-1.38	0.167	-.5287022	.091598
_cons	11.41203	2.028732	5.63	0.000	7.429801	15.39426

44 . outreg2 using "PS1_regression.doc", replace ctitle(Reg)
   PS1_regression.doc
   dir : seeout

45 .
46 . **iv) Estimates
47 .
48 . graph twoway (lfit cigs educ) (scatter cigs educ)

49 .
50 . **v)
51 .
52 . reg cigs educ, noconstant

Source	SS	df	MS		Number of obs	=	807
					F(1, 806)	=	283.28
Model	55301.1489	1	55301.1489		Prob > F	=	0.0000
Residual	157344.851	806	195.216937		R-squared	=	0.2601
					Adj R-squared	=	0.2591
Total	212646	807	263.501859		Root MSE	=	13.972

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ	.6447271	.0383061	16.83	0.000	.5695357	.7199186

53 . outreg2 using "PS1_regression_noconstant.doc", replace ctitle(Reg)
   PS1_regression_noconstant.doc
   dir : seeout

54 . twoway (lfit cigs educ) (lfit cigs educ, estopts(noconstant)) (scatter cigs educ)

55 .
56 . *8d)
57 . **i)
58 .
59 . gen age2 = age^2

60 . reg cigs educ age age2 white restaurn

Source	SS	df	MS		Number of obs	=	807
					F(5, 801)	=	8.65
Model	7772.46759	5	1554.49352		Prob > F	=	0.0000
Residual	143981.215	801	179.751829		R-squared	=	0.0512
					Adj R-squared	=	0.0453
Total	151753.683	806	188.280003		Root MSE	=	13.407

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ	-.4515013	.1615884	-2.79	0.005	-.768688	-.1343146
age	.8257641	.1544737	5.35	0.000	.5225431	1.128985
age2	-.009631	.0016817	-5.73	0.000	-.012932	-.00633
white	-.6237386	1.45611	-0.43	0.669	-3.481981	2.234504
restaurn	-2.796182	1.103552	-2.53	0.011	-4.962377	-.6299866
_cons	.6688335	3.706849	0.18	0.857	-6.607451	7.945118

61 . outreg2 using "PS1_regression2.doc", replace ctitle(Reg)
   PS1_regression2.doc
   dir : seeout

```
62 .
63 . **ii)
64 .
65 . reg cigs educ age age2 white restaurn
```

Source	SS	df	MS
Model	7772.46759	5	1554.49352
Residual	143981.215	801	179.751829
Total	151753.683	806	188.280003

Number of obs	=	807
F(5, 801)	=	8.65
Prob > F	=	0.0000
R-squared	=	0.0512
Adj R-squared	=	0.0453
Root MSE	=	13.407

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ	-.4515013	.1615884	-2.79	0.005	-.768688	-.1343146
age	.8257641	.1544737	5.35	0.000	.5225431	1.128985
age2	-.009631	.0016817	-5.73	0.000	-.012932	-.00633
white	-.6237386	1.45611	-0.43	0.669	-3.481981	2.234504
restaurn	-2.796182	1.103552	-2.53	0.011	-4.962377	-.6299866
_cons	.6688335	3.706849	0.18	0.857	-6.607451	7.945118

```
66 . mfx, varlist(age age2)

   Marginal effects after regress
        y  = Fitted values (predict)
           =  8.6864932
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
age	.8257641	.15447	5.35	0.000	.523001	1.12853	41.2379
age2	-.009631	.00168	-5.73	0.000	-.012927	-.006335	1990.14

```
67 .
68 . **iii)
69 . ***A)
70 . reg cigs educ age age2 white restaurn
```

Source	SS	df	MS
Model	7772.46759	5	1554.49352
Residual	143981.215	801	179.751829
Total	151753.683	806	188.280003

Number of obs	=	807
F(5, 801)	=	8.65
Prob > F	=	0.0000
R-squared	=	0.0512
Adj R-squared	=	0.0453
Root MSE	=	13.407

cigs	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ	-.4515013	.1615884	-2.79	0.005	-.768688	-.1343146
age	.8257641	.1544737	5.35	0.000	.5225431	1.128985
age2	-.009631	.0016817	-5.73	0.000	-.012932	-.00633
white	-.6237386	1.45611	-0.43	0.669	-3.481981	2.234504
restaurn	-2.796182	1.103552	-2.53	0.011	-4.962377	-.6299866
_cons	.6688335	3.706849	0.18	0.857	-6.607451	7.945118

```
71 . rvpplot age

72 .
```

```
73 . ***B)
74 . predict age_res, residuals

75 . gen age_res1=age_res[_n-1]
   (1 missing value generated)

76 . reg age_res age_res1
```

Source	SS	df	MS	Number of obs	=	806
				F(1, 804)	=	0.01
Model	2.57931919	1	2.57931919	Prob > F	=	0.9045
Residual	143869.779	804	178.942511	R-squared	=	0.0000
				Adj R-squared	=	-0.0012
Total	143872.358	805	178.723426	Root MSE	=	13.377

age_res	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
age_res1	-.0042332	.0352595	-0.12	0.904	-.0734448	.0649784
_cons	.0129003	.4711827	0.03	0.978	-.9119931	.9377938

```
77 .
78 . ***C)
79 . hist age_res, frequency normal width(1)
   (bin=86, start=-15.029084, width=1)

80 .
81 .
82 .
   end of do-file
```