

```
*****
* A Practical Guide to Using Panel Data
* Simonetta Longhi and Alita Nandi
* ISER, University of Essex
* Chapter 8
*****
```

```
-----
name: <unnamed>
log: C:\My Documents\Example_Chapter8.log
log type: text
opened on: 1 Sep 2014, 15:18:49
```

```
.
.
.
. * 8.1. Introduction
. *-----
.
. use "$datadir/rindresp", clear

. rename r* *

. quietly mvdecode _all, mv(-9/-1)

.
. * When "quietly" is added before the command, Stata
. * will run the command without showing the results of the command
.
. * 8.2. Continuous Variables
. * 8.2.1. Correlations Coefficients vs. Regression Coefficients
. *-----
.
. label var paygu "Usual gross pay per month"

.
. set more on

.
. correlate paygu age
(obs=7067)

-----+-----
      |      paygu      age
-----+-----
paygu |      1.0000
age   |      0.1492      1.0000

. scatter paygu age, scheme(s1manual)

. more

. twoway (scatter paygu age) (lfit paygu age), scheme(s1manual)

. more

.
. generate LnW = ln(paygu)
(7351 missing values generated)

.
. kdensity paygu

. more

. kdensity LnW

. more

. twoway (scatter LnW age) (lfit LnW age), scheme(s1manual)

.
```

```

. set more off

.
. generate age2 = age^2
(1 missing value generated)

.
. generate Female = 1 if sex == 2
(6973 missing values generated)

. replace Female = 0 if sex == 1
(6239 real changes made)

. label var Female "Dummy for women"

.
. generate Married = 1 if mastat == 1 | mastat == 2 | mastat == 7
(5197 missing values generated)

. replace Married = 0 if (mastat >= 3 & mastat <= 6) | (mastat >= 8 & mastat < .)
(5157 real changes made)

. label var Married "Whether married or cohabiting"

.
. * qfachi: highest academic qualification
. recode qfachi (-9/-1 = .) (1 = 2)
(qfachi: 424 changes made)

. tabulate qfachi, gen(Q)

```

	highest academic qualification	Freq.	Percent	Cum.
	1st degree	2,023	15.23	15.23
hnd,hnc,teaching	a level	918	6.91	22.14
	o level	2,768	20.84	42.98
	cse	3,391	25.53	68.50
	none of these	671	5.05	73.55
		3,513	26.45	100.00
Total		13,284	100.00	

```

. label var Q1 "1st degree or higher"

. label var Q2 "HND,HNC,teaching"

. label var Q3 "A level"

. label var Q4 "O level"

. label var Q5 "CSE"

. label var Q6 "none of these qualif"

```

```

.
. * region2: Government Office Regions
. tabulate region2, gen(R)

```

	government office region	Freq.	Percent	Cum.
	north east	361	2.52	2.52
	north west	1,044	7.28	9.80
yorkshire & humber	east midlands	827	5.77	15.57
	west midlands	722	5.04	20.61
	east of england	714	4.98	25.59
	london	839	5.85	31.44
	south east	617	4.30	35.74
	south west	1,168	8.15	43.89
	wales	812	5.66	49.56
	scotland	2,563	17.88	67.44
		2,462	17.17	84.61

northern ireland		2,206	15.39	100.00
-----+				
Total		14,335	100.00	

```
. label var R1 "North East"
. label var R2 "North West"
. label var R3 "Yorkshire & Humber"
. label var R4 "East Midlands"
. label var R5 "West Midlands"
. label var R6 "East of England"
. label var R7 "London"
. label var R8 "South East"
. label var R9 "South West"
. label var R10 "Wales"
. label var R11 "Scotland"
. label var R12 "Northern Ireland"
```

```
.
. save DatasetR.dta, replace
file DatasetR.dta saved
```

```
.
. regress LnW age
```

Source	SS	df	MS	Number of obs =	7067
-----+					
Model	151.535381	1	151.535381	F(1, 7065) =	233.95
Residual	4576.13267	7065	.647718708	Prob > F =	0.0000
-----+					
Total	4727.66805	7066	.66907275	R-squared =	0.0321
				Adj R-squared =	0.0319
				Root MSE =	.80481

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+					
age	.0111543	.0007293	15.30	0.000	.0097247 .0125838
_cons	6.748847	.0302577	223.05	0.000	6.689533 6.808161

```
. * Compare with:
. regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12
```

Source	SS	df	MS	Number of obs =	6933
-----+					
Model	1763.50556	20	88.1752779	F(20, 6912) =	217.90
Residual	2797.04324	6912	.40466482	Prob > F =	0.0000
-----+					
Total	4560.54879	6932	.657897979	R-squared =	0.3867
				Adj R-squared =	0.3849
				Root MSE =	.63613

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+					
age	.1277878	.0036589	34.93	0.000	.1206152 .1349604
age2	-.0014583	.0000442	-33.01	0.000	-.0015449 -.0013717
Female	-.5173127	.0153669	-33.66	0.000	-.5474366 -.4871888
Married	.0835274	.0188041	4.44	0.000	.0466656 .1203891
Q1	.7055081	.0287987	24.50	0.000	.6490539 .7619623
Q2	.5010076	.0358421	13.98	0.000	.4307461 .5712691
Q3	.2901391	.0280411	10.35	0.000	.2351699 .3451084
Q4	.1872705	.0272528	6.87	0.000	.1338467 .2406943
Q5	.1165528	.0396332	2.94	0.003	.0388596 .194246
R1	-.2786627	.0589121	-4.73	0.000	-.3941486 -.1631768
R2	-.3066754	.0458084	-6.69	0.000	-.3964739 -.2168769

R3	-.2909338	.0483127	-6.02	0.000	-.3856416	-.196226
R4	-.3418577	.0491452	-6.96	0.000	-.4381973	-.245518
R5	-.3572488	.0496795	-7.19	0.000	-.4546359	-.2598618
R6	-.258527	.0486662	-5.31	0.000	-.3539278	-.1631262
R8	-.2885082	.0444103	-6.50	0.000	-.3755661	-.2014503
R9	-.3170754	.0480772	-6.60	0.000	-.4113216	-.2228293
R10	-.3830289	.0413105	-9.27	0.000	-.4640102	-.3020475
R11	-.3182777	.0408353	-7.79	0.000	-.3983274	-.238228
R12	-.3309315	.0419792	-7.88	0.000	-.4132236	-.2486394
_cons	4.878964	.0780401	62.52	0.000	4.725981	5.031946

. * 8.2.2. Different Types of Standard Errors

. regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12, vce(robust)

Linear regression	Number of obs =	6933
	F(20, 6912) =	199.52
	Prob > F	= 0.0000
	R-squared	= 0.3867
	Root MSE	= .63613

LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1277878	.0042482	30.08	0.000	.1194601	.1361156
age2	-.0014583	.0000519	-28.11	0.000	-.00156	-.0013566
Female	-.5173127	.0151984	-34.04	0.000	-.5471062	-.4875193
Married	.0835274	.0177792	4.70	0.000	.0486747	.1183801
Q1	.7055081	.0295252	23.90	0.000	.6476296	.7633866
Q2	.5010076	.0362811	13.81	0.000	.4298855	.5721297
Q3	.2901391	.0289532	10.02	0.000	.2333821	.3468962
Q4	.1872705	.0287069	6.52	0.000	.1309963	.2435448
Q5	.1165528	.0373302	3.12	0.002	.0433741	.1897315
R1	-.2786627	.0536325	-5.20	0.000	-.3837988	-.1735266
R2	-.3066754	.0464689	-6.60	0.000	-.3977687	-.2155822
R3	-.2909338	.0466247	-6.24	0.000	-.3823325	-.1995351
R4	-.3418577	.0483738	-7.07	0.000	-.4366852	-.2470301
R5	-.3572488	.0509676	-7.01	0.000	-.457161	-.2573366
R6	-.258527	.0514016	-5.03	0.000	-.35929	-.157764
R8	-.2885082	.0464486	-6.21	0.000	-.3795616	-.1974547
R9	-.3170754	.0474077	-6.69	0.000	-.4100092	-.2241417
R10	-.3830289	.0417081	-9.18	0.000	-.4647895	-.3012682
R11	-.3182777	.040951	-7.77	0.000	-.3985543	-.2380012
R12	-.3309315	.041912	-7.90	0.000	-.4130919	-.2487711
_cons	4.878964	.089349	54.61	0.000	4.703812	5.054115

. * 8.2.3. Post-estimation Commands

```
. predict yhat_a, xb
(1250 missing values generated)

. label var yhat_a "Predictions (xb) on full sample"

. predict yhat_a_se, stdp
(1250 missing values generated)

. predict yhat_b if e(sample), xb
(7486 missing values generated)

. label var yhat_a "Predictions (xb) on estimation sample"

. predict yhat_b_se if e(sample), stdp
(7486 missing values generated)
```

```
. test Female
```

```
( 1) Female = 0
```

```
F( 1, 6912) = 1158.55
Prob > F = 0.0000
```

```
. test Q1 Q2 Q3 Q4 Q5
```

```
( 1) Q1 = 0
```

```
( 2) Q2 = 0
```

```
( 3) Q3 = 0
```

```
( 4) Q4 = 0
```

```
( 5) Q5 = 0
```

```
F( 5, 6912) = 190.02
Prob > F = 0.0000
```

```
.
. * Run the model again after including a dummy for part-time workers
. * jbf: employed full time
. tabulate jbf, gen(PT)
```

employed full time	Freq.	Percent	Cum.
full time: 30 hrs +	6,337	75.21	75.21
part time: lt 30 hrs	2,089	24.79	100.00
Total	8,426	100.00	

```
. rename PT2 PartTime
```

```
. regress LnW age age2 Female Married PartTime Q1-Q5 R1-R6 R8-R12, vce(robust)
```

```
Linear regression      Number of obs =    6880
                      F( 21, 6858) =    400.56
                      Prob > F      =    0.0000
                      R-squared      =    0.6244
                      Root MSE     =    .49619
```

LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.088841	.0034059	26.08	0.000	.0821643	.0955177
age2	-.0009707	.0000424	-22.88	0.000	-.0010538	-.0008875
Female	-.2256467	.0127961	-17.63	0.000	-.250731	-.2005623
Married	.1138963	.0140509	8.11	0.000	.0863523	.1414403
PartTime	-.987484	.0180455	-54.72	0.000	-1.022859	-.9521092
Q1	.6088579	.0238686	25.51	0.000	.5620681	.6556478
Q2	.4082738	.0292568	13.95	0.000	.3509214	.4656261
Q3	.2751816	.0230068	11.96	0.000	.2300811	.3202821
Q4	.160291	.022508	7.12	0.000	.1161683	.2044136
Q5	.0863405	.0293789	2.94	0.003	.0287487	.1439323
R1	-.2884608	.0431974	-6.68	0.000	-.3731412	-.2037805
R2	-.2902021	.0385291	-7.53	0.000	-.365731	-.2146731
R3	-.2645453	.0388755	-6.80	0.000	-.3407534	-.1883372
R4	-.2965224	.0401363	-7.39	0.000	-.3752019	-.2178428
R5	-.3299882	.0421835	-7.82	0.000	-.412681	-.2472954
R6	-.2236844	.0417228	-5.36	0.000	-.3054741	-.1418947
R8	-.2172875	.0379952	-5.72	0.000	-.29177	-.1428051
R9	-.2628452	.0387062	-6.79	0.000	-.3387213	-.186969
R10	-.3401631	.035284	-9.64	0.000	-.4093307	-.2709955
R11	-.2572789	.0346461	-7.43	0.000	-.3251961	-.1893618
R12	-.3076073	.0351798	-8.74	0.000	-.3765706	-.238644
_cons	5.655043	.0706139	80.08	0.000	5.516618	5.793468

```
.
.
. * 8.2.4. Diagnostics
. *-----
.
```

```
. regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12
```

Source	SS	df	MS	Number of obs =	6933
Model	1763.50556	20	88.1752779	F(20, 6912) =	217.90
Residual	2797.04324	6912	.40466482	Prob > F =	0.0000
				R-squared =	0.3867
				Adj R-squared =	0.3849
Total	4560.54879	6932	.657897979	Root MSE =	.63613

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.1277878	.0036589	34.93	0.000	.1206152 .1349604
age2	-.0014583	.0000442	-33.01	0.000	-.0015449 -.0013717
Female	-.5173127	.0153669	-33.66	0.000	-.5474366 -.4871888
Married	.0835274	.0188041	4.44	0.000	.0466656 .1203891
Q1	.7055081	.0287987	24.50	0.000	.6490539 .7619623
Q2	.5010076	.0358421	13.98	0.000	.4307461 .5712691
Q3	.2901391	.0280411	10.35	0.000	.2351699 .3451084
Q4	.1872705	.0272528	6.87	0.000	.1338467 .2406943
Q5	.1165528	.0396332	2.94	0.003	.0388596 .194246
R1	-.2786627	.0589121	-4.73	0.000	-.3941486 -.1631768
R2	-.3066754	.0458084	-6.69	0.000	-.3964739 -.2168769
R3	-.2909338	.0483127	-6.02	0.000	-.3856416 -.196226
R4	-.3418577	.0491452	-6.96	0.000	-.4381973 -.245518
R5	-.3572488	.0496795	-7.19	0.000	-.4546359 -.2598618
R6	-.258527	.0486662	-5.31	0.000	-.3539278 -.1631262
R8	-.2885082	.0444103	-6.50	0.000	-.3755661 -.2014503
R9	-.3170754	.0480772	-6.60	0.000	-.4113216 -.2228293
R10	-.3830289	.0413105	-9.27	0.000	-.4640102 -.3020475
R11	-.3182777	.0408353	-7.79	0.000	-.3983274 -.238228
R12	-.3309315	.0419792	-7.88	0.000	-.4132236 -.2486394
_cons	4.878964	.0780401	62.52	0.000	4.725981 5.031946

```
.
. * Heteroskedasticity
. set more on
```

```
. rvfplot, yline(0)
```

```
. more
```

```
. estat hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of LnW
```

```
chi2(1)          =    271.81
Prob > chi2      =    0.0000
```

```
.
. * Multicollinearity
. estat vif
```

Variable	VIF	1/VIF
age	39.32	0.025434
age2	38.10	0.026244
R11	4.18	0.239391
R10	3.94	0.253791
R12	3.62	0.276435
R8	2.82	0.354686
Q4	2.56	0.390217
R2	2.55	0.391940
Q3	2.53	0.395479
Q1	2.37	0.422075
R9	2.24	0.445704
R3	2.22	0.450460
R6	2.17	0.459784
R4	2.13	0.468803
R5	2.08	0.481658

R1	1.58	0.630958
Q2	1.57	0.636926
Q5	1.48	0.675560
Married	1.30	0.770387
Female	1.01	0.992633

Mean VIF	5.99
----------	------

```
. collin age age2 Female Married Q1-Q5 R1-R6 R8-R12
(obs=13169)
```

Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R- Squared
----------	-----	-------------	-----------	---------------

age	32.74	5.72	0.0305	0.9695
age2	32.52	5.70	0.0307	0.9693
Female	1.01	1.01	0.9898	0.0102
Married	1.30	1.14	0.7718	0.2282
Q1	1.58	1.26	0.6347	0.3653
Q2	1.24	1.11	0.8058	0.1942
Q3	1.75	1.32	0.5726	0.4274
Q4	1.72	1.31	0.5806	0.4194
Q5	1.25	1.12	0.7969	0.2031
R1	1.58	1.26	0.6327	0.3673
R2	2.59	1.61	0.3859	0.6141
R3	2.29	1.51	0.4361	0.5639
R4	2.13	1.46	0.4687	0.5313
R5	2.10	1.45	0.4753	0.5247
R6	2.30	1.52	0.4347	0.5653
R8	2.77	1.67	0.3604	0.6396
R9	2.28	1.51	0.4387	0.5613
R10	4.39	2.09	0.2280	0.7720
R11	4.31	2.08	0.2321	0.7679
R12	4.18	2.04	0.2394	0.7606

Mean VIF	5.30
----------	------

	Eigenval	Cond Index
--	----------	---------------

1	5.6064	1.0000
2	1.0871	2.2709
3	1.0479	2.3130
4	1.0384	2.3236
5	1.0167	2.3483
6	1.0033	2.3639
7	1.0004	2.3672
8	1.0002	2.3675
9	1.0001	2.3677
10	1.0000	2.3678
11	1.0000	2.3678
12	0.9872	2.3831
13	0.9685	2.4060
14	0.9565	2.4210
15	0.9204	2.4680
16	0.5262	3.2641
17	0.4411	3.5653
18	0.2789	4.4837
19	0.0995	7.5063
20	0.0185	17.4225
21	0.0028	45.0067

Condition Number 45.0067

Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)

Det(correlation matrix) 0.0015

```
.
. * Normality
. predict Resid, residuals
(7486 missing values generated)
```

```

. kdensity Resid, normal

. more

.
. mvtest norm Resid, stats(all)

Test for multivariate normality

Mardia mSkewness = .6036374      chi2(1) = 698.107      Prob>chi2 = 0.0000
Mardia mKurtosis = 4.703037      chi2(1) = 837.835      Prob>chi2 = 0.0000
Henze-Zirkler    = 45.98647      chi2(1) = 78.707      Prob>chi2 = 0.0000
Doornik-Hansen   =                chi2(2) = 555.507      Prob>chi2 = 0.0000

.
. set more off

.
. * 8.2.5. Interaction Terms
. *-----
. generate MarriedWoman = Married * Female
(770 missing values generated)

. regress LnW age age2 Female Married MarriedWoman Q1-Q5 R1-R6 R8-R12, vce(robust)

```

```

Linear regression                                Number of obs =      6933
                                                F( 21,  6911) =    197.69
                                                Prob > F       =    0.0000
                                                R-squared      =    0.3950
                                                Root MSE     =    .63186

```

	LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
	age	.1276517	.0042391	30.11	0.000	.1193417	.1359617
	age2	-.0014612	.0000517	-28.24	0.000	-.0015626	-.0013598
	Female	-.2948867	.029477	-10.00	0.000	-.3526708	-.2371026
	Married	.2621487	.0248542	10.55	0.000	.2134269	.3108706
MarriedWoman		-.3218283	.0342282	-9.40	0.000	-.388926	-.2547305
	Q1	.7027223	.0292402	24.03	0.000	.6454025	.7600421
	Q2	.4988589	.0360588	13.83	0.000	.4281726	.5695453
	Q3	.2826265	.0285748	9.89	0.000	.226611	.3386419
	Q4	.1857889	.0283883	6.54	0.000	.1301392	.2414387
	Q5	.1159745	.0370145	3.13	0.002	.0434146	.1885344
	R1	-.2761127	.0537536	-5.14	0.000	-.3814864	-.1707391
	R2	-.2988173	.0464111	-6.44	0.000	-.3897972	-.2078373
	R3	-.2868429	.0464893	-6.17	0.000	-.3779762	-.1957096
	R4	-.3388793	.048426	-7.00	0.000	-.4338092	-.2439494
	R5	-.3559591	.0508977	-6.99	0.000	-.4557343	-.2561839
	R6	-.2522411	.0510619	-4.94	0.000	-.3523382	-.1521441
	R8	-.2839727	.0463723	-6.12	0.000	-.3748766	-.1930688
	R9	-.3158565	.0473003	-6.68	0.000	-.4085796	-.2231334
	R10	-.38301	.0416297	-9.20	0.000	-.464617	-.301403
	R11	-.3159382	.0409452	-7.72	0.000	-.3962034	-.235673
	R12	-.3337361	.0419031	-7.96	0.000	-.415879	-.2515932
	_cons	4.764389	.089859	53.02	0.000	4.588238	4.94054

```

. test Married + MarriedWoman = 0

```

```

( 1) Married + MarriedWoman = 0

```

```

      F( 1,  6911) =    6.06
      Prob > F =    0.0138

```

```

.
. * Vs separate estimations: men
. regress LnW age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0, vce(robust)

```

```

Linear regression                                Number of obs =    3248
                                                F( 19,  3228) =    75.14
                                                Prob > F       =    0.0000

```


R-squared = 0.3997
Root MSE = .56676

LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1443334	.0057842	24.95	0.000	.1329922	.1556745
age2	-.0016376	.000069	-23.73	0.000	-.0017729	-.0015022
Married	.217188	.0246666	8.80	0.000	.1688242	.2655517
Q1	.5101071	.0377811	13.50	0.000	.4360296	.5841845
Q2	.407089	.0453442	8.98	0.000	.3181826	.4959954
Q3	.197006	.0365656	5.39	0.000	.1253118	.2687002
Q4	.1809222	.0365909	4.94	0.000	.1091785	.2526659
Q5	.1751289	.0455008	3.85	0.000	.0859155	.2643424
R1	-.3050504	.0726482	-4.20	0.000	-.4474917	-.1626091
R2	-.2574072	.0582356	-4.42	0.000	-.3715897	-.1432248
R3	-.2422732	.0600793	-4.03	0.000	-.3600706	-.1244759
R4	-.2927718	.0624387	-4.69	0.000	-.4151952	-.1703483
R5	-.3331775	.0662792	-5.03	0.000	-.4631311	-.203224
R6	-.1817815	.065793	-2.76	0.006	-.3107817	-.0527812
R8	-.2300818	.0598177	-3.85	0.000	-.3473663	-.1127973
R9	-.2411579	.0604966	-3.99	0.000	-.3597736	-.1225422
R10	-.3658288	.0547779	-6.68	0.000	-.4732318	-.2584258
R11	-.2802566	.0539071	-5.20	0.000	-.3859522	-.1745611
R12	-.3271627	.0540389	-6.05	0.000	-.4331167	-.2212087
_cons	4.477692	.1213288	36.91	0.000	4.239803	4.715582

. estimates store ResultsMen

. * and women

. regress LnW age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 1, vce(robust)

Linear regression

Number of obs = 3685
F(19, 3665) = 75.17
Prob > F = 0.0000
R-squared = 0.2884
Root MSE = .67667

LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1157469	.0061697	18.76	0.000	.1036505	.1278434
age2	-.0013265	.0000771	-17.21	0.000	-.0014776	-.0011755
Married	-.0327202	.0245923	-1.33	0.183	-.0809362	.0154959
Q1	.879351	.0440968	19.94	0.000	.7928943	.9658077
Q2	.5932097	.0565818	10.48	0.000	.4822748	.7041447
Q3	.3704825	.0442011	8.38	0.000	.2838214	.4571436
Q4	.212737	.0427821	4.97	0.000	.128858	.2966161
Q5	.0704396	.0573601	1.23	0.220	-.0420214	.1829006
R1	-.2508162	.0771073	-3.25	0.001	-.4019936	-.0996388
R2	-.3379547	.0709184	-4.77	0.000	-.4769982	-.1989112
R3	-.3247389	.0704872	-4.61	0.000	-.4629369	-.1865409
R4	-.3955709	.0727315	-5.44	0.000	-.5381691	-.2529728
R5	-.3825186	.0760007	-5.03	0.000	-.5315264	-.2335108
R6	-.3158734	.0764549	-4.13	0.000	-.4657717	-.165975
R8	-.3327406	.0694428	-4.79	0.000	-.468891	-.1965902
R9	-.3891743	.0721264	-5.40	0.000	-.5305862	-.2477624
R10	-.4083766	.0620517	-6.58	0.000	-.5300359	-.2867174
R11	-.3587405	.0610653	-5.87	0.000	-.4784658	-.2390153
R12	-.3467539	.0627197	-5.53	0.000	-.4697229	-.223785
_cons	4.651776	.1271073	36.60	0.000	4.402568	4.900984

. estimates store ResultsWomen

.
.
.
.
* 8.3 Estimating Weighted Regression Models
*-----

```

.
. * Add in psu and strata variables from rhhsamp using link variable rhid
. * The variable identifying households in the rhhsamp file is called rhid
. * while above we have renamed it hid
. generate double rhid = hid

```

```

. merge m:1 rhid using "$datadir\rhhsamp", keepusing(rhid rpsu rstrata)
(label rregion2 already defined)
(label rregion already defined)

```

Result	# of obs.
not matched	3,293
from master	0 (_merge==1)
from using	3,293 (_merge==2)
matched	14,419 (_merge==3)

```

. keep if _merge == 3
(3293 observations deleted)

```

```

. drop _merge

```

```

. rename rpsu psu

```

```

. rename rstrata strata

```

```

. regress LnW age age2 Female Married Q1-Q5

```

Source	SS	df	MS	Number of obs =	6985
Model	1730.16543	9	192.240604	F(9, 6975) =	470.39
Residual	2850.54501	6975	.408680288	Prob > F =	0.0000
				R-squared =	0.3777
				Adj R-squared =	0.3769
Total	4580.71044	6984	.655886375	Root MSE =	.63928

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.1270167	.0036603	34.70	0.000	.1198414 .134192
age2	-.0014477	.0000442	-32.75	0.000	-.0015344 -.0013611
Female	-.5163072	.015376	-33.58	0.000	-.5464488 -.4861655
Married	.0832343	.0187549	4.44	0.000	.046469 .1199995
Q1	.7241378	.0287271	25.21	0.000	.6678239 .7804518
Q2	.5051909	.0357219	14.14	0.000	.435165 .5752167
Q3	.2988562	.0280033	10.67	0.000	.2439613 .3537512
Q4	.1945445	.02723	7.14	0.000	.1411654 .2479235
Q5	.1258957	.0395816	3.18	0.001	.0483038 .2034877
_cons	4.573313	.0691988	66.09	0.000	4.437663 4.708964

```

. estimates store R_Regions

```

```

. regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12

```

Source	SS	df	MS	Number of obs =	6933
Model	1763.50556	20	88.1752779	F(20, 6912) =	217.90
Residual	2797.04324	6912	.40466482	Prob > F =	0.0000
				R-squared =	0.3867
				Adj R-squared =	0.3849
Total	4560.54879	6932	.657897979	Root MSE =	.63613

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.1277878	.0036589	34.93	0.000	.1206152 .1349604
age2	-.0014583	.0000442	-33.01	0.000	-.0015449 -.0013717
Female	-.5173127	.0153669	-33.66	0.000	-.5474366 -.4871888
Married	.0835274	.0188041	4.44	0.000	.0466656 .1203891
Q1	.7055081	.0287987	24.50	0.000	.6490539 .7619623

Q2	.5010076	.0358421	13.98	0.000	.4307461	.5712691
Q3	.2901391	.0280411	10.35	0.000	.2351699	.3451084
Q4	.1872705	.0272528	6.87	0.000	.1338467	.2406943
Q5	.1165528	.0396332	2.94	0.003	.0388596	.194246
R1	-.2786627	.0589121	-4.73	0.000	-.3941486	-.1631768
R2	-.3066754	.0458084	-6.69	0.000	-.3964739	-.2168769
R3	-.2909338	.0483127	-6.02	0.000	-.3856416	-.196226
R4	-.3418577	.0491452	-6.96	0.000	-.4381973	-.245518
R5	-.3572488	.0496795	-7.19	0.000	-.4546359	-.2598618
R6	-.258527	.0486662	-5.31	0.000	-.3539278	-.1631262
R8	-.2885082	.0444103	-6.50	0.000	-.3755661	-.2014503
R9	-.3170754	.0480772	-6.60	0.000	-.4113216	-.2228293
R10	-.3830289	.0413105	-9.27	0.000	-.4640102	-.3020475
R11	-.3182777	.0408353	-7.79	0.000	-.3983274	-.238228
R12	-.3309315	.0419792	-7.88	0.000	-.4132236	-.2486394
_cons	4.878964	.0780401	62.52	0.000	4.725981	5.031946

```

. estimates store R_NoWeights

```

```

. regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12 [pweight = xrwtuk1]
(sum of wgt is 7.7209e+03)

```

Linear regression	Number of obs =	6769
	F(20, 6748) =	112.82
	Prob > F =	0.0000
	R-squared =	0.3869
	Root MSE =	.6493

LnW	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1308828	.0056207	23.29	0.000	.1198645	.1419012
age2	-.0014913	.0000675	-22.11	0.000	-.0016236	-.0013591
Female	-.5496539	.0205501	-26.75	0.000	-.5899386	-.5093692
Married	.0417188	.0242714	1.72	0.086	-.0058608	.0892984
Q1	.7029947	.0408313	17.22	0.000	.6229524	.7830371
Q2	.4995007	.0522556	9.56	0.000	.3970632	.6019381
Q3	.3080371	.0396209	7.77	0.000	.2303676	.3857066
Q4	.1890128	.0396809	4.76	0.000	.1112256	.2667999
Q5	.1160016	.0479183	2.42	0.016	.0220667	.2099366
R1	-.2581424	.0570714	-4.52	0.000	-.3700203	-.1462645
R2	-.2920184	.0508405	-5.74	0.000	-.3916818	-.1923551
R3	-.276581	.0510785	-5.41	0.000	-.376711	-.176451
R4	-.3315176	.0532502	-6.23	0.000	-.4359047	-.2271304
R5	-.3444206	.0560619	-6.14	0.000	-.4543196	-.2345217
R6	-.2343718	.0575599	-4.07	0.000	-.3472074	-.1215363
R8	-.2476696	.0502791	-4.93	0.000	-.3462324	-.1491068
R9	-.2770463	.051823	-5.35	0.000	-.3786358	-.1754568
R10	-.3655642	.0461064	-7.93	0.000	-.4559473	-.275181
R11	-.2813341	.0446743	-6.30	0.000	-.3689099	-.1937583
R12	-.3047024	.0453409	-6.72	0.000	-.3935849	-.2158199
_cons	4.823145	.1169873	41.23	0.000	4.593813	5.052477

```

. estimates store R_Weights

```

```

. replace psu = hid if memorig == 7
psu was int now long
(2247 real changes made)

. svyset psu [pw=xrwtuk1], strata(strata)

    pweight: xrwtuk1
      VCE: linearized
Single unit: missing
  Strata 1: strata
    SU 1: psu
    FPC 1: <zero>

```

```
. svy: regress LnW age age2 Female Married Q1-Q5 R1-R6 R8-R12
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata	=	121	Number of obs	=	7449
Number of PSUs	=	1051	Population size	=	7720.8555
			Design df	=	930
			F(20, 911)	=	111.03
			Prob > F	=	0.0000
			R-squared	=	0.3869

LnW	Linearized		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
age	.1308828	.006282	20.83	0.000	.1185543	.1432114
age2	-.0014913	.0000744	-20.04	0.000	-.0016374	-.0013453
Female	-.5496539	.0210723	-26.08	0.000	-.5910087	-.5082991
Married	.0417188	.0228738	1.82	0.068	-.0031715	.0866091
Q1	.7029947	.0441149	15.94	0.000	.6164185	.789571
Q2	.4995007	.0520258	9.60	0.000	.3973992	.6016021
Q3	.3080371	.0409191	7.53	0.000	.2277326	.3883416
Q4	.1890128	.0416538	4.54	0.000	.1072664	.2707592
Q5	.1160016	.0516338	2.25	0.025	.0146693	.2173339
R1	-.2581424	.0602998	-4.28	0.000	-.3764818	-.139803
R2	-.2920184	.0439435	-6.65	0.000	-.3782584	-.2057785
R3	-.276581	.0476351	-5.81	0.000	-.3700658	-.1830962
R4	-.3315176	.0470253	-7.05	0.000	-.4238056	-.2392295
R5	-.3444206	.0545619	-6.31	0.000	-.4514994	-.2373419
R6	-.2343718	.0596534	-3.93	0.000	-.3514428	-.1173009
R8	-.2476696	.0467212	-5.30	0.000	-.3393608	-.1559783
R9	-.2770463	.0489909	-5.66	0.000	-.3731918	-.1809008
R10	-.3655642	.0419731	-8.71	0.000	-.4479371	-.2831912
R11	-.2813341	.0421597	-6.67	0.000	-.3640734	-.1985948
R12	-.3047024	.0415092	-7.34	0.000	-.386165	-.2232398
_cons	4.823145	.1268506	38.02	0.000	4.574198	5.072091

```
. estimates store R_SampleDesign
```

```
.
. * Save the results in a table *
. * See Chapter 14 *
. *****
.
. * Table 8.1
. quietly estout R_Regions R_NoWeights R_Weights R_SampleDesign ///
> using "WageModels.out", ///
> cells(b(star fmt(%9.3f)) se(par fmt(%9.3f))) ///
> style(tab) stats(r2 N, fmt(%9.3f %9.0g) ///
> labels(R2 Observations)) label collabels(, none) ///
> starlevels(+ 0.05 * 0.01) ///
> postfoot("St. err. in parenthesis; + Significant 5%, * Significant 1%") ///
> replace

.
.
. * 8.4. Binary Variables
. * 8.4.1. Linear and Non-linear models
. *-----
.
. * jbstat: current economic activity
. recode jbstat (1 2 = 1) (3/10 = 0), gen(Employed)
(13412 differences between jbstat and Employed)

. label define Employed 1 "Employed/Self-Employed" 2 "Unemployed or Inactive"

. label value Employed Employed

.
. regress Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0
```

Source	SS	df	MS	Number of obs =	5940
Model	558.396592	19	29.3892943	F(19, 5920) =	211.92
Residual	821.003408	5920	.138683008	Prob > F =	0.0000
				R-squared =	0.4048
				Adj R-squared =	0.4029
Total	1379.4	5939	.232261323	Root MSE =	.3724

Employed	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0458287	.001493	30.70	0.000	.0429018	.0487556
age2	-.0005592	.0000148	-37.87	0.000	-.0005882	-.0005303
Married	.1201092	.0120352	9.98	0.000	.0965158	.1437025
Q1	.1993653	.0166232	11.99	0.000	.1667778	.2319528
Q2	.1363423	.020626	6.61	0.000	.0959078	.1767767
Q3	.1071289	.0152612	7.02	0.000	.0772113	.1370464
Q4	.124841	.0149177	8.37	0.000	.0955968	.1540851
Q5	.1644242	.0247038	6.66	0.000	.1159958	.2128526
R1	-.0616644	.0387444	-1.59	0.112	-.1376176	.0142888
R2	-.053881	.0300772	-1.79	0.073	-.1128433	.0050813
R3	-.0265466	.0311654	-0.85	0.394	-.0876421	.0345489
R4	-.0352551	.0320938	-1.10	0.272	-.0981707	.0276606
R5	.0058044	.0327448	0.18	0.859	-.0583873	.0699962
R6	-.0237179	.0313097	-0.76	0.449	-.0850963	.0376605
R8	-.0001204	.0295816	-0.00	0.997	-.0581112	.0578703
R9	.0061724	.0313138	0.20	0.844	-.055214	.0675588
R10	-.0515056	.0269569	-1.91	0.056	-.104351	.0013399
R11	-.0296591	.0269214	-1.10	0.271	-.0824347	.0231166
R12	-.0797261	.0272793	-2.92	0.003	-.1332034	-.0262487
_cons	-.242146	.0411605	-5.88	0.000	-.3228355	-.1614565

. probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0

Iteration 0: log likelihood = -3903.5171
Iteration 1: log likelihood = -2374.8328
Iteration 2: log likelihood = -2338.7052
Iteration 3: log likelihood = -2338.5193
Iteration 4: log likelihood = -2338.5193

Probit regression	Number of obs =	5940
	LR chi2(19) =	3130.00
	Prob > chi2 =	0.0000
Log likelihood = -2338.5193	Pseudo R2 =	0.4009

Employed	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.2200136	.0075631	29.09	0.000	.2051902	.234837
age2	-.0027216	.0000835	-32.58	0.000	-.0028853	-.0025578
Married	.4931326	.051643	9.55	0.000	.3919141	.5943511
Q1	.7584157	.0735767	10.31	0.000	.614208	.9026235
Q2	.5049339	.0898357	5.62	0.000	.3288591	.6810087
Q3	.3788537	.0617057	6.14	0.000	.2579127	.4997948
Q4	.46473	.0616138	7.54	0.000	.3439692	.5854907
Q5	.545053	.1032919	5.28	0.000	.3426045	.7475015
R1	-.3251041	.1685243	-1.93	0.054	-.6554057	.0051976
R2	-.2565466	.1328258	-1.93	0.053	-.5168803	.0037872
R3	-.1615072	.1388417	-1.16	0.245	-.433632	.1106176
R4	-.207326	.140309	-1.48	0.140	-.4823266	.0676746
R5	.0904659	.1482148	0.61	0.542	-.2000298	.3809616
R6	-.0959101	.1397225	-0.69	0.492	-.3697611	.1779408
R8	.0515328	.1336144	0.39	0.700	-.2103467	.3134123
R9	.0260588	.1416574	0.18	0.854	-.2515846	.3037022
R10	-.2158928	.1208026	-1.79	0.074	-.4526616	.0208759
R11	-.1216355	.1211537	-1.00	0.315	-.3590923	.1158213
R12	-.3415943	.1217306	-2.81	0.005	-.5801819	-.1030068
_cons	-3.651224	.1887046	-19.35	0.000	-4.021079	-3.28137

Note: 16 failures and 0 successes completely determined.

. logit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0

```

Iteration 0: log likelihood = -3903.5171
Iteration 1: log likelihood = -2340.083
Iteration 2: log likelihood = -2325.0702
Iteration 3: log likelihood = -2325.0082
Iteration 4: log likelihood = -2325.0082

```

```

Logistic regression
Number of obs   =      5940
LR chi2(19)    =      3157.02
Prob > chi2     =      0.0000
Pseudo R2      =      0.4044

Log likelihood = -2325.0082

```

Employed	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.3930998	.014043	27.99	0.000	.3655761	.4206235
age2	-.0048742	.0001584	-30.77	0.000	-.0051847	-.0045637
Married	.8708715	.0930483	9.36	0.000	.6885002	1.053243
Q1	1.270692	.1327634	9.57	0.000	1.01048	1.530903
Q2	.862255	.1607684	5.36	0.000	.5471546	1.177355
Q3	.6347254	.1079846	5.88	0.000	.4230794	.8463713
Q4	.8366904	.1096673	7.63	0.000	.6217465	1.051634
Q5	.967299	.1865986	5.18	0.000	.6015725	1.333026
R1	-.5854061	.3007329	-1.95	0.052	-1.174832	.0040196
R2	-.4663927	.2381768	-1.96	0.050	-.9332106	.0004251
R3	-.2992171	.2485978	-1.20	0.229	-.7864598	.1880255
R4	-.3771497	.2528525	-1.49	0.136	-.8727315	.1184321
R5	.0918717	.2690435	0.34	0.733	-.4354439	.6191872
R6	-.148681	.2533863	-0.59	0.557	-.6453091	.3479471
R8	.0417137	.2411467	0.17	0.863	-.4309252	.5143526
R9	.0341559	.2561193	0.13	0.894	-.4678286	.5361404
R10	-.39651	.2177851	-1.82	0.069	-.8233609	.0303409
R11	-.22711	.2185459	-1.04	0.299	-.6554521	.2012321
R12	-.6153422	.2193408	-2.81	0.005	-1.045242	-.1854421
_cons	-6.496169	.3398631	-19.11	0.000	-7.162289	-5.83005

```

.
. * 8.4.2. Coefficients, Marginal Effects, Odd Ratios
. *-----
.
. probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0

```

```

Iteration 0: log likelihood = -3903.5171
Iteration 1: log likelihood = -2374.8328
Iteration 2: log likelihood = -2338.7052
Iteration 3: log likelihood = -2338.5193
Iteration 4: log likelihood = -2338.5193

```

```

Probit regression
Number of obs   =      5940
LR chi2(19)    =      3130.00
Prob > chi2     =      0.0000
Pseudo R2      =      0.4009

Log likelihood = -2338.5193

```

Employed	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.2200136	.0075631	29.09	0.000	.2051902	.234837
age2	-.0027216	.0000835	-32.58	0.000	-.0028853	-.0025578
Married	.4931326	.051643	9.55	0.000	.3919141	.5943511
Q1	.7584157	.0735767	10.31	0.000	.614208	.9026235
Q2	.5049339	.0898357	5.62	0.000	.3288591	.6810087
Q3	.3788537	.0617057	6.14	0.000	.2579127	.4997948
Q4	.46473	.0616138	7.54	0.000	.3439692	.5854907
Q5	.545053	.1032919	5.28	0.000	.3426045	.7475015
R1	-.3251041	.1685243	-1.93	0.054	-.6554057	.0051976
R2	-.2565466	.1328258	-1.93	0.053	-.5168803	.0037872
R3	-.1615072	.1388417	-1.16	0.245	-.433632	.1106176
R4	-.207326	.140309	-1.48	0.140	-.4823266	.0676746
R5	.0904659	.1482148	0.61	0.542	-.2000298	.3809616
R6	-.0959101	.1397225	-0.69	0.492	-.3697611	.1779408
R8	.0515328	.1336144	0.39	0.700	-.2103467	.3134123
R9	.0260588	.1416574	0.18	0.854	-.2515846	.3037022
R10	-.2158928	.1208026	-1.79	0.074	-.4526616	.0208759

Note: 16 failures and 0 successes completely determined.

```
> dydx(age age2 Married Q1 Q2 Q3 Q4 Q5 R1 R2 R3 R4 R5 R6 R8 R9 R10 R11 R12)
```

```
Expression      : Pr(Employed), predict()
dy/dx w.r.t.   : age age2 Married Q1 Q2 Q3 Q4 Q5 R1 R2 R3 R4 R5 R6 R8 R9 R10 R11 R12
```

Probit regression	Number of obs	=	5940
	LR chi2(19)	=	3130.00
	Prob > chi2	=	0.0000
Log likelihood = -2338.5193	Pseudo R2	=	0.4009

Employed	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.2200136	.0075631	29.09	0.000	.2051902	.234837
age2	-.0027216	.0000835	-32.58	0.000	-.0028853	-.0025578
1.Married	.4931326	.051643	9.55	0.000	.3919141	.5943511
1.Q1	.7584157	.0735767	10.31	0.000	.614208	.9026235
1.Q2	.5049339	.0898357	5.62	0.000	.3288591	.6810087
1.Q3	.3788537	.0617057	6.14	0.000	.2579127	.4997948
1.Q4	.46473	.0616138	7.54	0.000	.3439692	.5854907
1.Q5	.545053	.1032919	5.28	0.000	.3426045	.7475015
1.R1	-.3251041	.1685243	-1.93	0.054	-.6554057	.0051976
1.R2	-.2565466	.1328258	-1.93	0.053	-.5168803	.0037872
1.R3	-.1615072	.1388417	-1.16	0.245	-.433632	.1106176
1.R4	-.207326	.140309	-1.48	0.140	-.4823266	.0676746
1.R5	.0904659	.1482148	0.61	0.542	-.2000298	.3809616
1.R6	-.0959101	.1397225	-0.69	0.492	-.3697611	.1779408
1.R8	.0515328	.1336144	0.39	0.700	-.2103467	.3134123
1.R9	.0260588	.1416574	0.18	0.854	-.2515846	.3037022

1.R10	-.2158928	.1208026	-1.79	0.074	-.4526616	.0208759
1.R11	-.1216355	.1211537	-1.00	0.315	-.3590923	.1158213
1.R12	-.3415943	.1217306	-2.81	0.005	-.5801819	-.1030068
_cons	-3.651224	.1887046	-19.35	0.000	-4.021079	-3.28137

Note: 16 failures and 0 successes completely determined.

```
. margins, ///
>      dydx(age age2 Married Q1 Q2 Q3 Q4 Q5 R1 R2 R3 R4 R5 R6 R8 R9 R10 R11 R12)

Average marginal effects      Number of obs   =      5940
Model VCE      : OIM

Expression      : Pr(Employed), predict()
dy/dx w.r.t.    : age age2 1.Married 1.Q1 1.Q2 1.Q3 1.Q4 1.Q5 1.R1 1.R2 1.R3 1.R4 1.R5 1.R6 1.R8
                  1.R9 1.R10 1.R11 1.R12
```

		Delta-method				
	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0487101	.0013236	36.80	0.000	.0461159	.0513044
age2	-.0006025	.0000135	-44.55	0.000	-.0006291	-.000576
1.Married	.1161915	.0126079	9.22	0.000	.0914805	.1409025
1.Q1	.1533711	.0130198	11.78	0.000	.1278528	.1788895
1.Q2	.1019713	.0162217	6.29	0.000	.0701774	.1337652
1.Q3	.0809203	.0125951	6.42	0.000	.0562344	.1056063
1.Q4	.0976351	.0121324	8.05	0.000	.0738561	.1214141
1.Q5	.1086241	.0181684	5.98	0.000	.0730147	.1442336
1.R1	-.0761223	.0413966	-1.84	0.066	-.1572581	.0050135
1.R2	-.0591525	.0317442	-1.86	0.062	-.1213699	.003065
1.R3	-.0367343	.0323818	-1.13	0.257	-.1002015	.026733
1.R4	-.0475006	.0331625	-1.43	0.152	-.1124978	.0174967
1.R5	.0196944	.031706	0.62	0.534	-.0424483	.0818371
1.R6	-.0215857	.031944	-0.68	0.499	-.0841948	.0410233
1.R8	.0113116	.0290711	0.39	0.697	-.0456667	.0682899
1.R9	.0057428	.031073	0.18	0.853	-.0551592	.0666448
1.R10	-.0491829	.0282346	-1.74	0.082	-.1045217	.0061558
1.R11	-.0273432	.0276284	-0.99	0.322	-.0814938	.0268074
1.R12	-.0793679	.0294466	-2.70	0.007	-.1370823	-.0216536

Note: dy/dx for factor levels is the discrete change from the base level.

```
.
. logit Employed age age2 i.Married i.Q1-Q5 i.R1-R6 i.R8-R12 if Female == 0
```

```
Iteration 0:  log likelihood = -3903.5171
Iteration 1:  log likelihood = -2340.083
Iteration 2:  log likelihood = -2325.0702
Iteration 3:  log likelihood = -2325.0082
Iteration 4:  log likelihood = -2325.0082
```

Logistic regression	Number of obs	=	5940
	LR chi2(19)	=	3157.02
	Prob > chi2	=	0.0000
Log likelihood = -2325.0082	Pseudo R2	=	0.4044

	Employed	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age		.3930998	.014043	27.99	0.000	.3655761	.4206235
age2		-.0048742	.0001584	-30.77	0.000	-.0051847	-.0045637
1.Married		.8708715	.0930483	9.36	0.000	.6885002	1.053243
1.Q1		1.270692	.1327634	9.57	0.000	1.01048	1.530903
1.Q2		.862255	.1607684	5.36	0.000	.5471546	1.177355
1.Q3		.6347254	.1079846	5.88	0.000	.4230794	.8463713
1.Q4		.8366904	.1096673	7.63	0.000	.6217465	1.051634
1.Q5		.967299	.1865986	5.18	0.000	.6015725	1.333026
1.R1		-.5854061	.3007329	-1.95	0.052	-1.174832	.0040196
1.R2		-.4663927	.2381768	-1.96	0.050	-.9332106	.0004251
1.R3		-.2992171	.2485978	-1.20	0.229	-.7864598	.1880255
1.R4		-.3771497	.2528525	-1.49	0.136	-.8727315	.1184321
1.R5		.0918717	.2690435	0.34	0.733	-.4354439	.6191872


```
. margins, ///
>          dydx(age age2 Married Q1 Q2 Q3 Q4 Q5 R1 R2 R3 R4 R5 R6 R8 R9 R10 R11 R12)

Average marginal effects          Number of obs   =          5940
Model VCE      : OIM

Expression      : Pr(Employed), predict()
dy/dx w.r.t.    : age age2 1.Married 1.Q1 1.Q2 1.Q3 1.Q4 1.Q5 1.R1 1.R2 1.R3 1.R4 1.R5 1.R6 1.R8
                  1.R9 1.R10 1.R11 1.R12
```

	Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
age	.0486419	.0012579	38.67	0.000	.0461765	.0511073
age2	-.0006031	.0000129	-46.72	0.000	-.0006284	-.0005778
1.Married	.114972	.0126901	9.06	0.000	.0901	.1398441
1.Q1	.1444504	.0132563	10.90	0.000	.1184685	.1704323
1.Q2	.0977044	.0163436	5.98	0.000	.0656715	.1297372
1.Q3	.0759591	.0123862	6.13	0.000	.0516826	.1002357
1.Q4	.0979937	.0119651	8.19	0.000	.0745426	.1214447
1.Q5	.1076635	.0182892	5.89	0.000	.0718173	.1435096
1.R1	-.0768672	.041598	-1.85	0.065	-.1583978	.0046635
1.R2	-.0602797	.0320087	-1.88	0.060	-.1230156	.0024562
1.R3	-.0381171	.0325441	-1.17	0.242	-.1019023	.0256681
1.R4	-.0483994	.0335534	-1.44	0.149	-.1141629	.0173641
1.R5	.0112575	.0326401	0.34	0.730	-.0527159	.0752309
1.R6	-.0186722	.0322821	-0.58	0.563	-.081944	.0445996
1.R8	.0051411	.0296008	0.17	0.862	-.0528754	.0631575
1.R9	.0042117	.0314701	0.13	0.894	-.0574686	.065892
1.R10	-.050578	.0285574	-1.77	0.077	-.1065494	.0053935
1.R11	-.0285845	.0279517	-1.02	0.306	-.0833687	.0261998
1.R12	-.0800822	.0298078	-2.69	0.007	-.1385044	-.02166

Note: dy/dx for factor levels is the discrete change from the base level.

```
. logit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0, or
```

```
Iteration 0:    log likelihood = -3903.5171
Iteration 1:    log likelihood = -2340.083
Iteration 2:    log likelihood = -2325.0702
Iteration 3:    log likelihood = -2325.0082
Iteration 4:    log likelihood = -2325.0082
```

Logistic regression	Number of obs	=	5940
	LR chi2(19)	=	3157.02
	Prob > chi2	=	0.0000
Log likelihood = -2325.0082	Pseudo R2	=	0.4044

Employed	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	1.481566	.0208056	27.99	0.000	1.441344	1.522911
age2	.9951377	.0001577	-30.77	0.000	.9948287	.9954467
Married	2.388992	.2222915	9.36	0.000	1.990728	2.866933
Q1	3.563317	.4730782	9.57	0.000	2.74692	4.622351
Q2	2.368496	.3807793	5.36	0.000	1.728328	3.245779
Q3	1.886504	.2037134	5.88	0.000	1.526656	2.331172
Q4	2.308713	.2531903	7.63	0.000	1.862178	2.862325
Q5	2.630829	.490909	5.18	0.000	1.824986	3.792501
R1	.5568797	.167472	-1.95	0.052	.3088709	1.004028
R2	.6272609	.149399	-1.96	0.050	.393289	1.000425
R3	.7413984	.18431	-1.20	0.229	.4554543	1.206864

R4	.6858134	.1734096	-1.49	0.136	.4178088	1.12573
R5	1.096224	.294932	0.34	0.733	.6469774	1.857418
R6	.861844	.2183795	-0.59	0.557	.5245004	1.416157
R8	1.042596	.2514186	0.17	0.863	.6499075	1.672555
R9	1.034746	.2650184	0.13	0.894	.6263608	1.709397
R10	.6726636	.1464961	-1.82	0.069	.4389539	1.030806
R11	.7968331	.1741446	-1.04	0.299	.5192073	1.222909
R12	.5404559	.118544	-2.81	0.005	.3516066	.8307369
_cons	.0015092	.0005129	-19.11	0.000	.0007753	.0029379

. * 8.4.3. Different Types of Standard Errors and Other Post-estimation commands

. probit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0, vce(robust)

Iteration 0: log pseudolikelihood = -3903.5171
Iteration 1: log pseudolikelihood = -2374.8328
Iteration 2: log pseudolikelihood = -2338.7052
Iteration 3: log pseudolikelihood = -2338.5193
Iteration 4: log pseudolikelihood = -2338.5193

Probit regression	Number of obs	=	5940
	Wald chi2(19)	=	1229.89
	Prob > chi2	=	0.0000
Log pseudolikelihood = -2338.5193	Pseudo R2	=	0.4009

Employed	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.2200136	.009073	24.25	0.000	.2022309	.2377963
age2	-.0027216	.0001047	-26.01	0.000	-.0029267	-.0025165
Married	.4931326	.0528681	9.33	0.000	.3895131	.5967521
Q1	.7584157	.0753342	10.07	0.000	.6107634	.906068
Q2	.5049339	.0875708	5.77	0.000	.3332984	.6765694
Q3	.3788537	.061789	6.13	0.000	.2577495	.499958
Q4	.46473	.0636752	7.30	0.000	.3399288	.5895311
Q5	.545053	.10748	5.07	0.000	.3343961	.7557099
R1	-.3251041	.1659511	-1.96	0.050	-.6503621	.000154
R2	-.2565466	.1319917	-1.94	0.052	-.5152455	.0021524
R3	-.1615072	.1357756	-1.19	0.234	-.4276226	.1046081
R4	-.207326	.1417011	-1.46	0.143	-.485055	.0704031
R5	.0904659	.1582161	0.57	0.567	-.2196319	.4005637
R6	-.0959101	.1405627	-0.68	0.495	-.371408	.1795878
R8	.0515328	.135174	0.38	0.703	-.2134034	.3164689
R9	.0260588	.143673	0.18	0.856	-.2555351	.3076527
R10	-.2158928	.1200073	-1.80	0.072	-.4511028	.0193171
R11	-.1216355	.1197628	-1.02	0.310	-.3563663	.1130952
R12	-.3415943	.1209514	-2.82	0.005	-.5786546	-.104534
_cons	-3.651224	.2006922	-18.19	0.000	-4.044574	-3.257875

Note: 16 failures and 0 successes completely determined.

. logit Employed age age2 Married Q1-Q5 R1-R6 R8-R12 if Female == 0, vce(robust)

Iteration 0: log pseudolikelihood = -3903.5171
Iteration 1: log pseudolikelihood = -2340.083
Iteration 2: log pseudolikelihood = -2325.0702
Iteration 3: log pseudolikelihood = -2325.0082
Iteration 4: log pseudolikelihood = -2325.0082

Logistic regression	Number of obs	=	5940
	Wald chi2(19)	=	1267.41
	Prob > chi2	=	0.0000
Log pseudolikelihood = -2325.0082	Pseudo R2	=	0.4044

Employed	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.3930998	.0150316	26.15	0.000	.3636384	.4225612

age2	-.0048742	.0001721	-28.33	0.000	-.0052114	-.0045369
Married	.8708715	.0953178	9.14	0.000	.6840519	1.057691
Q1	1.270692	.1308689	9.71	0.000	1.014194	1.52719
Q2	.862255	.1576797	5.47	0.000	.5532084	1.171302
Q3	.6347254	.1089535	5.83	0.000	.4211805	.8482703
Q4	.8366904	.1137077	7.36	0.000	.6138274	1.059553
Q5	.967299	.197858	4.89	0.000	.5795044	1.355094
R1	-.5854061	.2994373	-1.96	0.051	-1.172292	.0014802
R2	-.4663927	.2377597	-1.96	0.050	-.9323932	-.0003923
R3	-.2992171	.2462149	-1.22	0.224	-.7817894	.1833551
R4	-.3771497	.2609185	-1.45	0.148	-.8885405	.1342411
R5	.0918717	.2771202	0.33	0.740	-.4512739	.6350173
R6	-.148681	.2580487	-0.58	0.564	-.6544472	.3570851
R8	.0417137	.2445754	0.17	0.865	-.4376453	.5210727
R9	.0341559	.2596905	0.13	0.895	-.4748282	.54314
R10	-.39651	.2192277	-1.81	0.071	-.8261883	.0331683
R11	-.22711	.2196934	-1.03	0.301	-.6577012	.2034812
R12	-.6153422	.2214674	-2.78	0.005	-1.04941	-.181274
_cons	-6.496169	.3457602	-18.79	0.000	-7.173847	-5.818492

```

.
. * 8.5. Multiple Outcomes
. * 8.5.1. Ordered Outcomes
. -----
.
. generate Unemployed = 1 if jbstat == 3
(13923 missing values generated)

. replace Unemployed = 0 if jbstat == 1 | jbstat == 2
(8181 real changes made)

.
. * lfsato: satisfaction with: life overall
. tab lfsato Unemployed, miss

```

satisfaction with:	Unemployed			Total
life overall	0	1	.	
-----+-----+-----+-----+-----				
not satisfied at all	33	17	108	158
2	85	25	147	257
3	348	63	351	762
not satis/dissat	983	88	688	1,759
5	2,700	128	1,374	4,202
6	2,893	93	1,606	4,592
completely satisfied	670	49	968	1,687
.	469	33	500	1,002
-----+-----+-----+-----+-----				
Total	8,181	496	5,742	14,419

```

.
. oprobit lfsato i.Unemployed i.Female age age2 i.Married i.Q1-Q5 i.R1-R6 i.R8-R12

```

```

Iteration 0: log likelihood = -10995.744
Iteration 1: log likelihood = -10829.266
Iteration 2: log likelihood = -10829.236
Iteration 3: log likelihood = -10829.236

```

```

Ordered probit regression              Number of obs   =       7539
                                      LR chi2(21)        =       333.02
                                      Prob > chi2         =       0.0000
Log likelihood = -10829.236           Pseudo R2        =       0.0151

```

lfsato	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.Unemployed	-.4828452	.0541581	-8.92	0.000	-.5889931	-.3766974
1.Female	.0171765	.0242933	0.71	0.480	-.0304375	.0647904
age	-.068527	.0059152	-11.58	0.000	-.0801206	-.0569334
age2	.0008033	.0000698	11.51	0.000	.0006665	.0009401
1.Married	.3405748	.0295841	11.51	0.000	.2825909	.3985587
1.Q1	-.0130264	.0442507	-0.29	0.768	-.0997562	.0737034

/cut3	-4.596357	.2364244	-5.05974	-4.132973
/cut4	-3.358247	.2331987	-3.815308	-2.901186
/cut5	-1.752493	.2311702	-2.205579	-1.299408
/cut6	.4595112	.2312727	.006225	.9127974

```
. regress lfsato Unemployed Female age age2 Married Q1-Q5 R1-R6 R8-R12
```

Source	SS	df	MS	Number of obs =	7539
Model	419.258833	21	19.9647063	F(21, 7517) =	17.19
Residual	8730.33156	7517	1.16141168	Prob > F =	0.0000
				R-squared =	0.0458
				Adj R-squared =	0.0432
Total	9149.5904	7538	1.21379549	Root MSE =	1.0777

lfsato	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Unemployed	-.5810668	.0558431	-10.41	0.000	-.690535 - .4715987
Female	.0081627	.0249518	0.33	0.744	-.0407499 .0570753
age	-.0651185	.0060173	-10.82	0.000	-.076914 - .0533229
age2	.0007597	.0000709	10.71	0.000	.0006207 .0008987
Married	.3631935	.0303262	11.98	0.000	.3037458 .4226413
Q1	.0227467	.0453889	0.50	0.616	-.0662283 .1117217
Q2	.0505857	.0572905	0.88	0.377	-.0617197 .1628912
Q3	-.0638055	.0441259	-1.45	0.148	-.1503046 .0226936
Q4	-.0413002	.0426692	-0.97	0.333	-.1249438 .0423433
Q5	.0552687	.0620582	0.89	0.373	-.0663827 .1769201
R1	-.0043919	.096706	-0.05	0.964	-.1939628 .185179
R2	.0397403	.0740046	0.54	0.591	-.1053293 .18481
R3	.0425633	.0774255	0.55	0.583	-.1092124 .194339
R4	-.0352317	.0788457	-0.45	0.655	-.1897914 .119328
R5	-.0799175	.0799338	-1.00	0.317	-.2366101 .0767751
R6	.0054773	.0773569	0.07	0.944	-.1461639 .1571186
R8	-.1016436	.0715905	-1.42	0.156	-.2419811 .0386938
R9	.0202057	.0770375	0.26	0.793	-.1308092 .1712207
R10	.0247946	.0663524	0.37	0.709	-.1052747 .1548639
R11	.0016312	.0658346	0.02	0.980	-.127423 .1306853
R12	.1548269	.0692573	2.24	0.025	.0190631 .2905906
_cons	6.26321	.1328302	47.15	0.000	6.002825 6.523594

```
. * 8.5.2. Unordered Outcomes
```

```
. *-----
```

```
. recode jbstat (1 2 = 1) (3 = 2) (4/10 = 3), gen(LMS)
(13412 differences between jbstat and LMS)
```

```
. label var LMS "Labour Market Status"
```

```
. label define LMS 1 "Has Job" 2 "Unemployed" 3 "Inactive"
```

```
. label values LMS LMS
```

```
. mprobit LMS i.Female age age2 i.Married i.Q1-Q5 i.R1-R6 i.R8-R12, baseoutcome(1)
```

```
Iteration 0: log likelihood = -7081.9916
Iteration 1: log likelihood = -7053.7331
Iteration 2: log likelihood = -7049.0935
Iteration 3: log likelihood = -7049.0699
Iteration 4: log likelihood = -7049.0699
```

Multinomial probit regression	Number of obs =	13168
	Wald chi2(40) =	3712.82
Log likelihood = -7049.0699	Prob > chi2 =	0.0000

LMS	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Has_Job	(base outcome)				

LnW	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.046388	.0073095	6.35	0.000	.0320557	.0607202
age2	-.0005002	.0000907	-5.52	0.000	-.000678	-.0003225
Married	.0192148	.0196193	0.98	0.327	-.0192541	.0576837
PartTime	-.8878366	.0211297	-42.02	0.000	-.9292671	-.8464062
Q1	.7766111	.0350035	22.19	0.000	.7079773	.8452449
Q2	.4979393	.0455754	10.93	0.000	.4085763	.5873023
Q3	.4191105	.0362539	11.56	0.000	.348025	.490196
Q4	.2296972	.0345773	6.64	0.000	.1618991	.2974953
Q5	.1030482	.0469546	2.19	0.028	.010981	.1951154
R1	-.2620804	.0630697	-4.16	0.000	-.3857457	-.1384151
R2	-.2914039	.0588672	-4.95	0.000	-.4068289	-.1759788
R3	-.2600921	.059621	-4.36	0.000	-.3769952	-.1431889
R4	-.3089352	.0612434	-5.04	0.000	-.4290194	-.1888509
R5	-.3379509	.0630519	-5.36	0.000	-.4615813	-.2143205
R6	-.2849778	.0617957	-4.61	0.000	-.406145	-.1638106
R8	-.217186	.0576624	-3.77	0.000	-.3302488	-.1041233
R9	-.2992388	.05913	-5.06	0.000	-.4151792	-.1832985
R10	-.3480776	.0524864	-6.63	0.000	-.4509914	-.2451637
R11	-.2463581	.0513369	-4.80	0.000	-.347018	-.1456981
R12	-.3125936	.0533044	-5.86	0.000	-.4171114	-.2080758
_cons	6.268379	.1489524	42.08	0.000	5.976317	6.56044

```
.
. heckman LnW age age2 Married PartTime Q1-Q5 R1-R6 R8-R12 ///
> if Female == 1 & age >= 23 & age < 60, ///
> select(Works = age age2 Married Q1-Q5 R1-R6 R8-R12 nchild) ///
> first mills(MR) vce(robust)
```

```
Iteration 0: log likelihood = -2383.8457
Iteration 1: log likelihood = -2111.997
Iteration 2: log likelihood = -2110.4567
Iteration 3: log likelihood = -2110.4567
```

```
Probit regression                                Number of obs   =      4070
                                                LR chi2(20)    =      546.78
                                                Prob > chi2     =      0.0000
Log likelihood = -2110.4567                    Pseudo R2      =      0.1147
```

Works	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.1340117	.0193797	6.92	0.000	.0960282	.1719952
age2	-.0018167	.0002366	-7.68	0.000	-.0022804	-.001353
Married	.2602291	.0516983	5.03	0.000	.1589022	.361556
Q1	1.077933	.0769023	14.02	0.000	.9272069	1.228658
Q2	.7523163	.0976412	7.70	0.000	.5609432	.9436895
Q3	.7591271	.0728852	10.42	0.000	.6162747	.9019795
Q4	.7123348	.066091	10.78	0.000	.5827988	.8418707
Q5	.3778668	.0971732	3.89	0.000	.1874109	.5683228
R1	.0190715	.1764556	0.11	0.914	-.326775	.3649181
R2	-.0081888	.1390256	-0.06	0.953	-.2806739	.2642962
R3	.0538256	.1467358	0.37	0.714	-.2337713	.3414225
R4	-.0809601	.1449699	-0.56	0.577	-.3650958	.2031756
R5	.1043767	.1545791	0.68	0.500	-.1985927	.4073461
R6	-.0290801	.1452352	-0.20	0.841	-.3137359	.2555758
R8	.2025496	.1387449	1.46	0.144	-.0693854	.4744846
R9	.110228	.149	0.74	0.459	-.1818065	.4022626
R10	-.0574524	.1248853	-0.46	0.645	-.3022231	.1873183
R11	.0583668	.1254117	0.47	0.642	-.1874357	.3041693
R12	-.1199183	.1248881	-0.96	0.337	-.3646944	.1248578
nchild	-.3066468	.0244229	-12.56	0.000	-.3545147	-.2587788
_cons	-2.188768	.3886447	-5.63	0.000	-2.950497	-1.427038

```
Iteration 0: log pseudolikelihood = -4141.6855
Iteration 1: log pseudolikelihood = -4141.0161
Iteration 2: log pseudolikelihood = -4141.001
Iteration 3: log pseudolikelihood = -4141.0009
```

```
Heckman selection model                        Number of obs   =      4070
```

```

(regression model with sample selection)      Censored obs      =      1109
                                              Uncensored obs    =      2961

                                              Wald chi2(20)      =      1964.13
                                              Prob > chi2        =      0.0000

Log pseudolikelihood = -4141.001

```

		Robust				
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
LnW						
age	.0403529	.0086761	4.65	0.000	.0233481	.0573577
age2	-.0004184	.0001098	-3.81	0.000	-.0006336	-.0002032
Married	.0024247	.0226891	0.11	0.915	-.042045	.0468945
PartTime	-.8655154	.0261085	-33.15	0.000	-.9166872	-.8143436
Q1	.6578638	.0775116	8.49	0.000	.5059439	.8097836
Q2	.4070711	.0708434	5.75	0.000	.2682205	.5459217
Q3	.3259123	.0665929	4.89	0.000	.1953927	.4564319
Q4	.1424577	.0645076	2.21	0.027	.0160252	.2688903
Q5	.0500665	.0579052	0.86	0.387	-.0634256	.1635587
R1	-.2640085	.064835	-4.07	0.000	-.3910828	-.1369342
R2	-.2902065	.0599824	-4.84	0.000	-.4077699	-.1726431
R3	-.267454	.0608267	-4.40	0.000	-.3866721	-.1482359
R4	-.2974514	.0627542	-4.74	0.000	-.4204474	-.1744555
R5	-.3463974	.064679	-5.36	0.000	-.473166	-.2196288
R6	-.2837482	.0628418	-4.52	0.000	-.406916	-.1605805
R8	-.236092	.0597249	-3.95	0.000	-.3531508	-.1190333
R9	-.3094822	.0606722	-5.10	0.000	-.4283976	-.1905668
R10	-.3404277	.0536754	-6.34	0.000	-.4456296	-.2352258
R11	-.2519995	.0524406	-4.81	0.000	-.3547811	-.1492179
R12	-.2943966	.0547654	-5.38	0.000	-.4017349	-.1870583
_cons	6.551172	.2358671	27.77	0.000	6.088881	7.013463
Works						
age	.1331478	.0193818	6.87	0.000	.0951602	.1711355
age2	-.0017987	.00024	-7.50	0.000	-.002269	-.0013283
Married	.2626056	.0504373	5.21	0.000	.1637503	.3614609
Q1	1.057794	.0814626	12.99	0.000	.8981305	1.217458
Q2	.7456624	.0967417	7.71	0.000	.5560521	.9352727
Q3	.7535347	.0738803	10.20	0.000	.608732	.8983375
Q4	.7120935	.0667925	10.66	0.000	.5811826	.8430043
Q5	.3771014	.0979621	3.85	0.000	.1850992	.5691036
R1	.0031586	.1745308	0.02	0.986	-.3389155	.3452326
R2	-.0098349	.1418904	-0.07	0.945	-.2879349	.2682651
R3	.0506498	.1472185	0.34	0.731	-.2378932	.3391928
R4	-.0931455	.1483827	-0.63	0.530	-.3839703	.1976793
R5	.0815787	.1554687	0.52	0.600	-.2231344	.3862918
R6	-.0388129	.1461277	-0.27	0.791	-.325218	.2475922
R8	.2012568	.1429861	1.41	0.159	-.0789907	.4815043
R9	.1001348	.1551209	0.65	0.519	-.2038965	.4041661
R10	-.0642644	.1285866	-0.50	0.617	-.3162894	.1877606
R11	.044923	.1307084	0.34	0.731	-.2112608	.3011068
R12	-.1190685	.127693	-0.93	0.351	-.3693422	.1312051
nchild	-.301542	.0315292	-9.56	0.000	-.363338	-.239746
_cons	-2.182475	.3803974	-5.74	0.000	-2.928041	-1.43691
/athrho	-.4847439	.2963721	-1.64	0.102	-1.065623	.0961349
/lnsigma	-.6879872	.0559823	-12.29	0.000	-.7977106	-.5782639
rho	-.4500348	.2363475			-.7878063	.0958398
sigma	.5025866	.028136			.4503589	.5608713
lambda	-.2261815	.1308381			-.4826193	.0302564

Wald test of indep. eqns. (rho = 0): chi2(1) = 2.68 Prob > chi2 = 0.1019

```

. log close
  name: <unnamed>
  log: C:\My Documents\Example_Chapter8.log
  log type: text
  closed on: 1 Sep 2014, 15:28:22

```