

PAF 573
Advanced Regression
Linear Probability/Probit/Logit/Logistic

```
log: C:\wvs 508b\handouts\linear probability_probit_logit_logistic handout.log
log type: text
opened on: 22 Mar 2004, 12:11:46
```

```
. use "C:\PUAF611\past semesters\Fall 2000\datasets\bcs002.dta", clear
(Breast Cancer treatment, 1990)
```

```
. describe
```

```
Contains data from C:\PUAF611\past semesters\Fall 2000\datasets\bcs002.dta
  obs:          10,130          Breast Cancer treatment, 1990
  vars:           32          13 May 1997 12:28
  size:         445,720 (99.9% of memory free)
```

```
-----
```

variable name	storage type	display format	value label	variable label
county	byte	%9.0g		County Code
hosp	int	%9.0g		Hospital Identifier
age	byte	%9.0g		Age
race	byte	%9.0g	race	Race
lenstay	byte	%9.0g		Length of stay
bcs	byte	%8.0g	bcs	mastectomy or breast conserving
hcount	int	%9.0g		Number of patients per hospital
advstage	byte	%8.0g		=1 if advanced stage cancer
forprof	byte	%8.0g		=1 if for profit hospital
hosptype	byte	%8.0g	hosptype	type of hospital~
insure	byte	%8.0g	insure	insurance type
managed	byte	%9.0g		
smurban	byte	%9.0g		
medurban	byte	%9.0g		
lrgurban	byte	%9.0g		
teaching	byte	%9.0g		
rural	byte	%9.0g		
othrhosp	byte	%9.0g		
white	byte	%9.0g		
black	byte	%9.0g		
hispanic	byte	%9.0g		
native	byte	%9.0g		
asian	byte	%9.0g		
other	byte	%9.0g		
unknown	byte	%9.0g		
medicare	byte	%9.0g		
medical	byte	%9.0g		
indemnit	byte	%9.0g		
hmoppp	byte	%9.0g		
indigent	byte	%9.0g		
poor	float	%9.0g		
age2	float	%9.0g		

```
-----
```

```
Sorted by:
```

```
. tab insure
```

insurance type	Freq.	Percent	Cum.
Medicare	4354	42.98	42.98
Medi-Cal	393	3.88	46.86
indemnit	2393	23.62	70.48
HMO/PPP	2913	28.76	99.24
indigent	77	0.76	100.00
Total	10130	100.00	

LINEAR PROBABILITY MODEL

```
. reg bcs indigent hmoppp medical medicare
```

Source	SS	df	MS	Number of obs =	10130
Model	11.2214012	4	2.8053503	F(4, 10125) =	19.76
Residual	1437.76735	10125	.142001713	Prob > F =	0.0000
Total	1448.98875	10129	.143053485	R-squared =	0.0077
				Adj R-squared =	0.0074
				Root MSE =	.37683

bcs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
indigent	-.0932807	.0436293	-2.14	0.033	-.1788029 -.0077586
hmoppp	-.0391481	.0103965	-3.77	0.000	-.0595274 -.0187688
medical	-.0755682	.0205102	-3.68	0.000	-.1157722 -.0353641
medicare	-.0821311	.0095893	-8.56	0.000	-.100928 -.0633342
_cons	.2231509	.0077033	28.97	0.000	.2080509 .2382508

```
. test indigent=medical
```

```
( 1) indigent - medical = 0.0
```

```
F( 1, 10125) = 0.14
Prob > F = 0.7061
```

```
. reg bcs indigent hmoppp medical medicare advstage age age2 lenstay hispanic native
black asian other hcount forprof managed smurban medurban teaching othrhop
```

Source	SS	df	MS	Number of obs =	10130
Model	72.3899905	20	3.61949952	F(20, 10109) =	26.58
Residual	1376.59876	10109	.136175562	Prob > F =	0.0000
				R-squared =	0.0500
				Adj R-squared =	0.0481
Total	1448.98875	10129	.143053485	Root MSE =	.36902

bcs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
indigent	-.061983	.0429692	-1.44	0.149	-.1462111 .0222451
hmoppp	-.0459342	.011324	-4.06	0.000	-.0681315 -.023737
medical	-.0379912	.0205952	-1.84	0.065	-.0783619 .0023795
medicare	-.024989	.0131077	-1.91	0.057	-.0506827 .0007046
advstage	-.0283164	.0084757	-3.34	0.001	-.0449305 -.0117023
age	-.0025818	.0020294	-1.27	0.203	-.0065599 .0013962
age2	6.45e-06	.000017	0.38	0.705	-.0000269 .0000398
lenstay	-.0346009	.0019176	-18.04	0.000	-.0383597 -.0308421
hispanic	-.0412965	.0160877	-2.57	0.010	-.0728317 -.0097614
native	.123448	.1508134	0.82	0.413	-.1721762 .4190722
black	.0294138	.0171552	1.71	0.086	-.0042139 .0630415
asian	-.0488998	.0183412	-2.67	0.008	-.0848521 -.0129474
other	.0538253	.0471273	1.14	0.253	-.0385535 .1462042
hcount	9.99e-06	9.16e-06	1.09	0.275	-7.97e-06 .000028
forprof	-.0441405	.0226005	-1.95	0.051	-.088442 .0001609
managed	.0237007	.0111503	2.13	0.034	.0018438 .0455575
smurban	-.039427	.0152942	-2.58	0.010	-.0694067 -.0094473
medurban	-.0320758	.0108886	-2.95	0.003	-.0534195 -.010732
teaching	.05687	.0152792	3.72	0.000	.0269198 .0868203
othrhosp	-.0138592	.0181558	-0.76	0.445	-.0494481 .0217297
_cons	.4355108	.0612657	7.11	0.000	.3154178 .5556038

```
. test indigent=medical
```

```
( 1) indigent - medical = 0.0
```

```
F( 1, 10109) = 0.27
Prob > F = 0.6031
```

PROBIT MODEL

```
. probit bcs indigent hmoppp medical medicare advstage age age2 lenstay hispanic native  
black asian other hcount forprof managed smurban medurban teaching othr hosp
```

```
Iteration 0: log likelihood = -4665.2263  
Iteration 1: log likelihood = -4380.3792  
Iteration 2: log likelihood = -4345.0183  
Iteration 3: log likelihood = -4341.625  
Iteration 4: log likelihood = -4341.1904  
Iteration 5: log likelihood = -4341.1278  
Iteration 6: log likelihood = -4341.1186  
Iteration 7: log likelihood = -4341.1172
```

```
Probit estimates                               Number of obs   =      10130  
                                                LR chi2(20)    =      648.22  
                                                Prob > chi2    =      0.0000  
Log likelihood = -4341.1172                    Pseudo R2      =      0.0695
```

	bcs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
indigent		-.2410281	.194153	-1.24	0.214	-.621561 .1395049
hmoppp		-.1765814	.0455165	-3.88	0.000	-.265792 -.0873708
medical		-.1461697	.0883348	-1.65	0.098	-.3193027 .0269633
medicare		-.0912412	.0542939	-1.68	0.093	-.1976554 .015173
advstage		-.1188912	.036061	-3.30	0.001	-.1895695 -.048213
age		-.0066527	.008525	-0.78	0.435	-.0233614 .0100559
age2		-2.93e-06	.000073	-0.04	0.968	-.0001461 .0001402
lenstay		-.2102985	.0105749	-19.89	0.000	-.231025 -.1895721
hispanic		-.1706313	.0711283	-2.40	0.016	-.3100403 -.0312224
native		.4121886	.5497467	0.75	0.453	-.6652952 1.489672
black		.1114601	.0696623	1.60	0.110	-.0250756 .2479958
asian		-.2072345	.0793331	-2.61	0.009	-.3627245 -.0517446
other		.1872141	.1854932	1.01	0.313	-.1763459 .5507742
hcount		.0000405	.000038	1.06	0.287	-.0000341 .000115
forprof		-.1920936	.1003661	-1.91	0.056	-.3888075 .0046203
managed		.1116872	.0453233	2.46	0.014	.0228551 .2005192
smurban		-.188417	.0674752	-2.79	0.005	-.3206661 -.056168
medurban		-.1347226	.0466304	-2.89	0.004	-.2261165 -.0433287
teaching		.2190142	.060584	3.62	0.000	.1002716 .3377567
othrhosp		-.0582697	.0756097	-0.77	0.441	-.2064619 .0899226
_cons		.1224664	.2525146	0.48	0.628	-.3724531 .617386

note: 5 failures and 0 successes completely determined.

```
. test indigent=medical
```

```
( 1) indigent - medical = 0.0  
  
      chi2( 1) =      0.21  
      Prob > chi2 =      0.6486
```

Wald Test

```
. test advstage
```

```
( 1) advstage = 0.0  
  
      chi2( 1) =     10.87  
      Prob > chi2 =      0.0010
```

Likelihood Ratio Test

. test smurban medurban

- (1) smurban = 0.0
- (2) medurban = 0.0

chi2(2) = 12.98
 Prob > chi2 = 0.0015

. lrtest, saving(0)

. quietly probit bcs indigent hmoppp medical medicare advstage age age2 lenstay
 hispanic native black asian other hcount forprof managed teaching othrhop

. lrtest

Probit: likelihood-ratio test

chi2(2) = 13.62
 Prob > chi2 = 0.0011

Assessing Goodness of Fit

. lstat

Probit model for bcs

Classified	----- True -----		Total
	D	~D	
+	3	0	3
-	1749	8378	10127
Total	1752	8378	10130

Classified + if predicted Pr(D) >= .5
 True D defined as bcs ~= 0

Sensitivity	Pr(+ D)	0.17%
Specificity	Pr(- ~D)	100.00%
Positive predictive value	Pr(D +)	100.00%
Negative predictive value	Pr(~D -)	82.73%
False + rate for true ~D	Pr(+ ~D)	0.00%
False - rate for true D	Pr(- D)	99.83%
False + rate for classified +	Pr(~D +)	0.00%
False - rate for classified -	Pr(D -)	17.27%
Correctly classified		82.73%

Marginal Effects

```
. dprobit bcs indigent hmoppp medical medicare advstage age age2 lenstay hispanic
native black asian other hcount forprof managed smurban medurban teaching othrhosp
```

```
Iteration 0: log likelihood = -4665.2263
Iteration 1: log likelihood = -4380.3792
Iteration 2: log likelihood = -4345.0183
Iteration 3: log likelihood = -4341.625
Iteration 4: log likelihood = -4341.1904
Iteration 5: log likelihood = -4341.1278
Iteration 6: log likelihood = -4341.1186
Iteration 7: log likelihood = -4341.1172
```

Probit estimates

```
Number of obs = 10130
LR chi2(20) = 648.22
Prob > chi2 = 0.0000
Pseudo R2 = 0.0695
```

Log likelihood = -4341.1172

	bcs	dF/dx	Std. Err.	z	P> z	x-bar	[95% C.I.]
indigent*	-.0501625	.0350142	-1.24	0.214	.007601	-.118789	.018464
hmoppp*	-.0401621	.0099443	-3.88	0.000	.287562	-.059653	-.020672
medical*	-.0321936	.0180331	-1.65	0.098	.038796	-.067538	.003151
medicare*	-.0214346	.0126719	-1.68	0.093	.429812	-.046271	.003402
advstage*	-.0272763	.0080144	-3.30	0.001	.254294	-.042984	-.011568
age	-.0015732	.0020162	-0.78	0.435	61.3816	-.005525	.002379
age2	-6.93e-07	.0000173	-0.04	0.968	3967.39	-.000035	.000033
lenstay	-.0497286	.0024111	-19.89	0.000	2.9074	-.054454	-.045003
hispanic*	-.0372568	.0142511	-2.40	0.016	.058045	-.065188	-.009325
native*	.1175088	.1819624	0.75	0.453	.000592	-.239131	.474149
black*	.0277078	.0181584	1.60	0.110	.050148	-.007882	.063298
asian*	-.0443021	.0151829	-2.61	0.009	.043139	-.07406	-.014544
other*	.0484435	.0520957	1.01	0.313	.00612	-.053662	.150549
hcount	9.57e-06	8.99e-06	1.06	0.287	792.888	-8.1e-06	.000027
forprof*	-.0412483	.0193965	-1.91	0.056	.028233	-.079265	-.003232
managed*	.0273267	.0114556	2.46	0.014	.196249	.004874	.049779
smurban*	-.0409284	.0133657	-2.79	0.005	.074531	-.067125	-.014732
medurban*	-.0304105	.0100249	-2.89	0.004	.169102	-.050059	-.010762
teaching*	.0566877	.0170161	3.62	0.000	.076111	.023337	.090039
othrhosp*	-.0134062	.0169125	-0.77	0.441	.045607	-.046554	.019742
obs. P	.1729516						
pred. P	.1532128	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| are the test of the underlying coefficient being 0

Logit Model

```
. logit bcs indigent hmoppp medical medicare advstage age age2 lenstay hispanic native  
black asian other hcount forprof managed smurban medurban teaching othrhop
```

```
Iteration 0: log likelihood = -4665.2263  
Iteration 1: log likelihood = -4365.4207  
Iteration 2: log likelihood = -4281.1727  
Iteration 3: log likelihood = -4271.9859  
Iteration 4: log likelihood = -4271.7087
```

```
Logit estimates                                     Number of obs   =    10130  
                                                    LR chi2(20)    =    787.04  
                                                    Prob > chi2    =    0.0000  
Log likelihood = -4271.7087                       Pseudo R2      =    0.0844
```

bcs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
indigent	-.3333221	.3572713	-0.93	0.351	-1.033561 .3669167
hmoppp	-.3309386	.0806196	-4.10	0.000	-.4889502 -.172927
medical	-.2050203	.1604589	-1.28	0.201	-.519514 .1094733
medicare	-.1612885	.0983613	-1.64	0.101	-.3540731 .0314961
advstage	-.1816619	.0656322	-2.77	0.006	-.3102986 -.0530251
age	-.0124948	.0155986	-0.80	0.423	-.0430674 .0180778
age2	5.42e-06	.0001355	0.04	0.968	-.0002601 .000271
lenstay	-.5670141	.0272677	-20.79	0.000	-.6204577 -.5135704
hispanic	-.306348	.1322859	-2.32	0.021	-.5656237 -.0470723
native	.7610925	.9408878	0.81	0.419	-1.083014 2.605199
black	.1843885	.124308	1.48	0.138	-.0592507 .4280276
asian	-.3603975	.1448963	-2.49	0.013	-.6443891 -.0764059
other	.3082378	.3270226	0.94	0.346	-.3327149 .9491904
hcount	.0000837	.0000685	1.22	0.222	-.0000505 .0002179
forprof	-.3180383	.1863302	-1.71	0.088	-.6832387 .0471621
managed	.2027776	.0807886	2.51	0.012	.0444348 .3611205
smurban	-.3506863	.1252373	-2.80	0.005	-.5961469 -.1052256
medurban	-.2752252	.0855872	-3.22	0.001	-.4429731 -.1074773
teaching	.4269572	.1061179	4.02	0.000	.21897 .6349445
othrhosp	-.1397266	.1363955	-1.02	0.306	-.4070568 .1276036
_cons	.7435592	.4580439	1.62	0.105	-.1541903 1.641309
