

Models In Epidemiology And Biostatistics

Gordon Hilton Fick

Session 5 : Example

*** Data is in chds.dta

. gen bwtg=bwt*453.6

. gen lbw=(bwtg<3000)

. gen msmo=(mnocig>0)

. gen g=gestwks

. gen sg=msmo*g

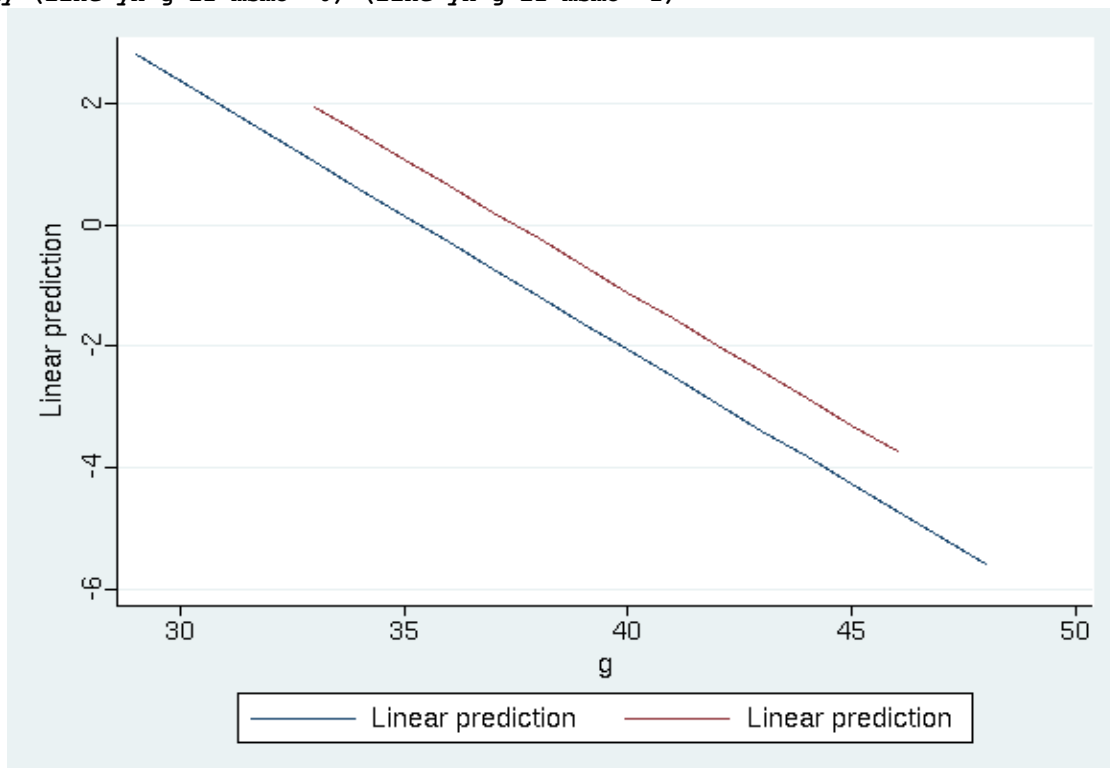
. logit lbw msmo g sg

Logistic regression	Number of obs	=	680
	LR chi2(3)	=	89.97
	Prob > chi2	=	0.0000
Log likelihood = -307.39691	Pseudo R2	=	0.1277

lbw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
msmo	.7405227	4.878673	0.15	0.879	-8.821501	10.30255
g	-.441808	.0914342	-4.83	0.000	-.6210158	-.2626002
sg	.0052407	.1245069	0.04	0.966	-.2387883	.2492698
_cons	15.61609	3.582078	4.36	0.000	8.595343	22.63683

. predict yh,xb

. twoway (line yh g if msmo==0) (line yh g if msmo==1)



```
. logit lbw msmo g
```

Logistic regression

```
Number of obs   =      680
LR chi2(2)       =      89.97
Prob > chi2      =      0.0000
Pseudo R2       =      0.1277
```

Log likelihood = -307.3978

lbw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
msmo	.9456984	.2041954	4.63	0.000	.5454827	1.345914
g	-.4389871	.0620676	-7.07	0.000	-.5606373	-.3173369
_cons	15.50567	2.434436	6.37	0.000	10.73426	20.27708

```
. predict yh2,xb
```

```
. twoway (line yh2 g if msmo==0) (line yh2 g if msmo==1)
```

```
. cc lbw msmo
```

			Proportion	
	Exposed	Unexposed	Total	Exposed
-----+-----+-----+-----+-----				
Cases	92	53	145	0.6345
Controls	207	328	535	0.3869
-----+-----+-----+-----+-----				
Total	299	381	680	0.4397
	Point estimate		[95% Conf. Interval]	
	-----+-----+-----+-----+-----			
Odds ratio	2.750524		1.850485	4.10555 (exact)
Attr. frac. ex.	.6364329		.459601	.7564273 (exact)
Attr. frac. pop	.4038057			
	+-----+-----+-----+-----+-----			
	chi2(1) =		28.38	Pr>chi2 = 0.0000

```
. logit lbw msmo
```

Logistic regression

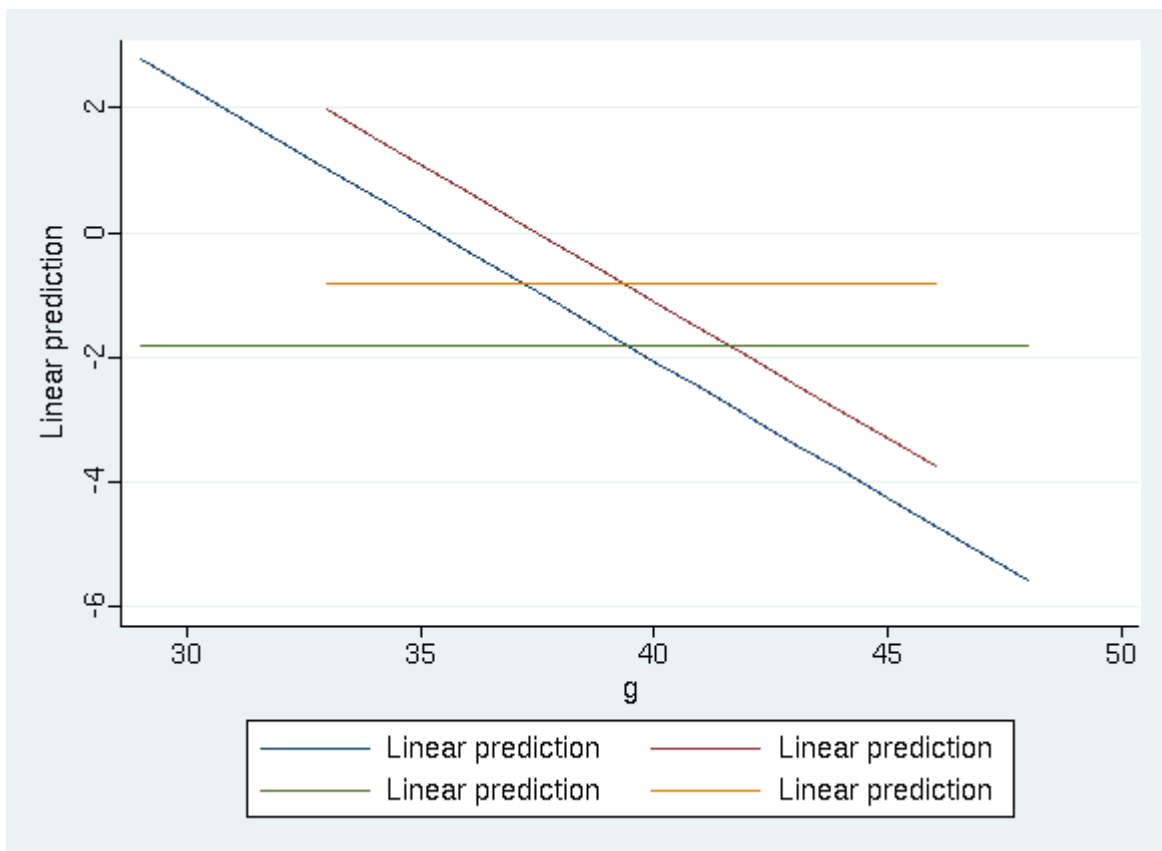
```
Number of obs   =      680
LR chi2(1)      =      28.31
Prob > chi2     =      0.0000
Pseudo R2      =      0.0402
```

Log likelihood = -338.22792

lbw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
msmo	1.011791	.1939515	5.22	0.000	.6316536	1.391929
_cons	-1.822722	.1480428	-12.31	0.000	-2.11288	-1.532563

```
. predict yh3,xb
```

```
. twoway (line yh2 g if msmo==0) (line yh2 g if msmo==1) (line yh3 g if msmo==0) (line yh3 g if msmo==1)
```



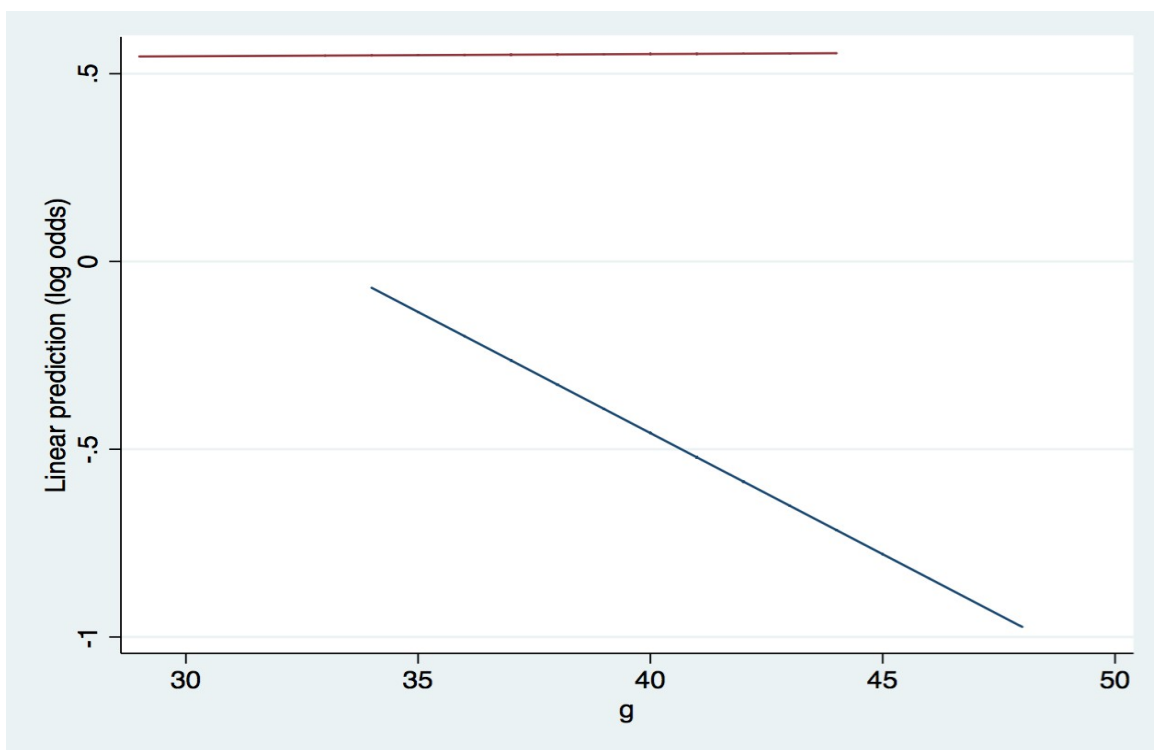
Interpret all the estimates with explanations and identify which lines go with which equations.

Let us now consider models for the log odds of maternal smoking. We can view this project as case-control which is, perhaps, more reasonable.

```
. gen lg=lbw*g
. logit msmo lbw g lg
```

Logistic regression	Number of obs	=	680
	LR chi2(3)	=	29.77
	Prob > chi2	=	0.0000
Log likelihood = -451.4987	Pseudo R2	=	0.0319

msmo	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbw	-1.593873	3.792401	-0.42	0.674	-9.026842	5.839096
g	-.064497	.0536866	-1.20	0.230	-.1697207	.0407268
lg	.0650805	.0968997	0.67	0.502	-.1248395	.2550004
_cons	2.12281	2.150835	0.99	0.324	-2.092749	6.33837



```
. logit msmo lbw g
```

Logistic regression

```
Number of obs    =      680
LR chi2(2)       =      29.32
Prob > chi2      =      0.0000
Pseudo R2       =      0.0314
```

Log likelihood = -451.72305

msmo	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbw	.9506019	.2029549	4.68	0.000	.5528177	1.348386
g	-.0447939	.0447577	-1.00	0.317	-.1325175	.0429297
_cons	1.333993	1.794381	0.74	0.457	-2.182928	4.850915

```
. logit msmo lbw
```

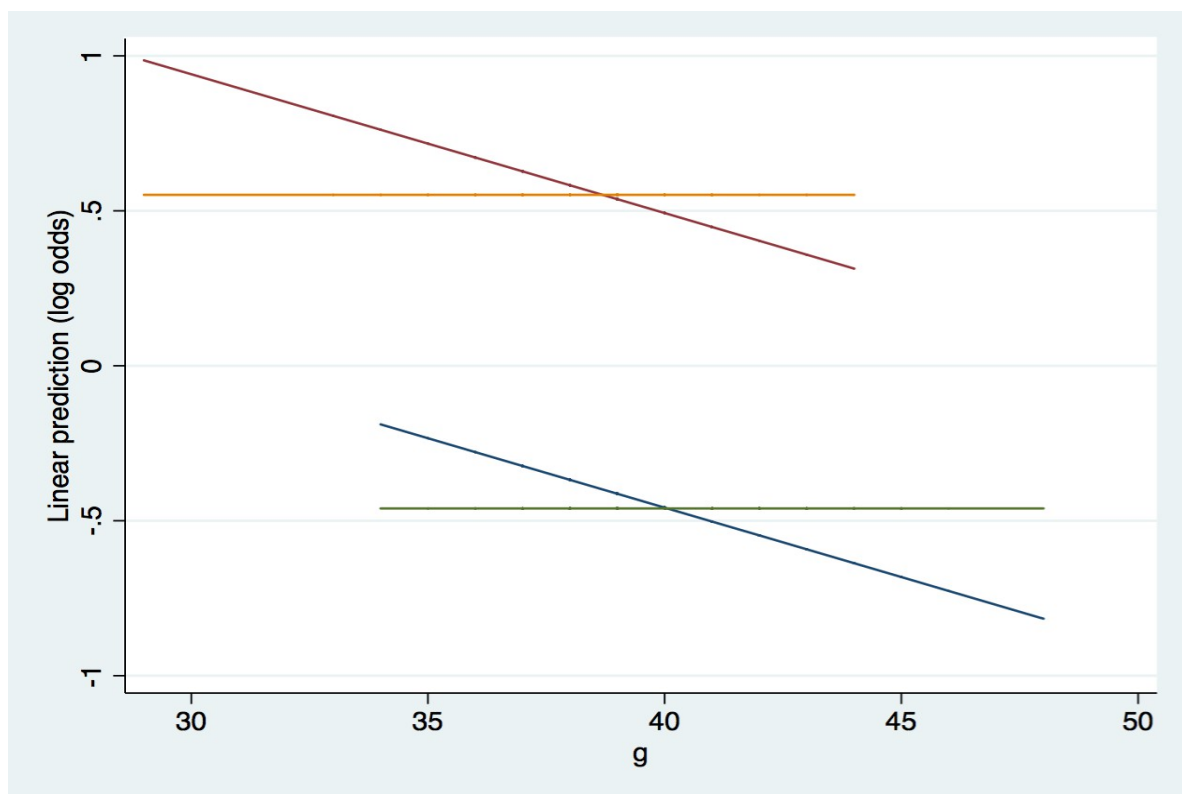
Logistic regression

```
Number of obs    =      680
LR chi2(1)       =      28.31
Prob > chi2      =      0.0000
Pseudo R2       =      0.0304
```

Log likelihood = -452.22783

msmo	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbw	1.011791	.1939515	5.22	0.000	.6316535	1.391929
_cons	-.4602948	.0887677	-5.19	0.000	-.6342762	-.2863134

Notice the models with log odds of low birthweight are not the same as the models with log of odds of maternal smoking. Parts are the same. Which parts? Why? Parts are different. Which parts? Why?



```
. summ g,d
```

g				

	Percentiles	Smallest		
1%	34	29		
5%	37	33		
10%	38	33	Obs	680
25%	39	34	Sum of Wgt.	680
50%	40		Mean	39.77059
		Largest	Std. Dev.	1.875433
75%	41	45		
90%	42	45	Variance	3.517249
95%	43	46	Skewness	-.2181004
99%	45	48	Kurtosis	5.980773

Consider centring the gestational age at 40 weeks

```
. gen gc=g-40
```

```
. gen sgc=msmo*gc
```

```
. logit lbw msmo gc sgc
```

```
Logistic regression               Number of obs   =       680
                                LR chi2(3)        =       89.97
                                Prob > chi2       =       0.0000
Log likelihood = -307.39691       Pseudo R2      =       0.1277
```

lbw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
msmo	.9501521	.2300303	4.13	0.000	.499301	1.401003
gc	-.441808	.0914343	-4.83	0.000	-.6210159	-.2626001
sgc	.0052407	.1245069	0.04	0.966	-.2387884	.2492698
_cons	-2.056233	.1739747	-11.82	0.000	-2.397217	-1.715248

Notice that b_0 and b_1 change from the uncentred model while b_2 and b_3 do not change. We have that b_0 and b_1 are specific to age=0 with the uncentred model and that b_0 and b_1 are specific to age=40 with the centred model while b_2 and b_3 refer to slopes and are hence not changed by centring. Write out the two models. Solve the first in terms of the second and solve the second in terms of the first.

Now lets consider changing the scale of gestational age from weeks to days.

```
. gen gcd=gc*7
```

```
. gen sgcd=msmo*gcd
```

```
. logit lbw msmo gcd sgcd
```

```
Iteration 0:  log likelihood =  -352.384
Iteration 1:  log likelihood = -310.44763
Iteration 2:  log likelihood = -307.42039
Iteration 3:  log likelihood = -307.39691
Iteration 4:  log likelihood = -307.39691
```

```
Logistic regression               Number of obs   =       680
                                LR chi2(3)        =       89.97
                                Prob > chi2       =       0.0000
Log likelihood = -307.39691       Pseudo R2      =       0.1277
```

lbw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
msmo	.9501521	.2300303	4.13	0.000	.499301	1.401003
gcd	-.0631154	.013062	-4.83	0.000	-.0887166	-.0375143
sgcd	.0007487	.0177867	0.04	0.966	-.0341126	.03561
_cons	-2.056233	.1739747	-11.82	0.000	-2.397217	-1.715248

Notice that the two slope estimates (b_2 and b_3) are changed but the age specific estimates (b_0 and b_1) are not changed. The 'new' b_2 and b_3 are now estimates of the rates of change of low birthweight per day of gestational age. They are the old estimates divided by 7.