Quantitative Data Analysis II

SOC 781

Review binary & ordinal outcomes

Today we will…

- review logit and ologit

Odds ratios: logit

- (0) = not happy (1) = happy

```
logit hap_dic c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
```

hap_dic	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9594691	.0038904	-10.20	0.000	.9518743	.9671246
c.age#c.age	1.000402	.0000404	9.96	0.000	1.000323	1.000481
1.female	1.098327	.0280566	3.67	0.000	1.044691	1.154717
1.nonwhite	.6423927	.0185629	-15.32	0.000	.6070212	.6798253
1.married	2.79002	.0766205	37.36	0.000	2.643816	2.944308
educat						
1	1.509675	.0488828	12.72	0.000	1.416843	1.608589
2	2.169886	.0680512	24.70	0.000	2.040525	2.307449
_cons	7.418151	.6972194	21.32	0.000	6.170106	8.91864

- Interpretation?

Odds ratios: ologit

- (1) = not too happy (2) = pretty happy (3) = very happy

```
ologit hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
```

hap	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9680552	.0026094	-12.04	0.000	.9629543	.9731831
c.age#c.age	1.000362	.0000268	13.50	0.000	1.00031	1.000415
1.female	1.121912	.0182835	7.06	0.000	1.086644	1.158326
1.nonwhite	.7349438	.0155689	-14.54	0.000	.7050541	.7661005
1.married	2.680555	.046998	56.24	0.000	2.590005	2.77427
educat						
1	1.231638	.0282539	9.08	0.000	1.177488	1.288278
2	1.578385	.033964	21.21	0.000	1.513201	1.646377

- What major assumption should be met to justify interpretation of these odds ratios?

- Let’s assume it’s met…Interpretation?

Linear regression vs. ologit

- Substantively compare results

```
reg hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0
```

hap	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0100741	.0008349	-12.07	0.000	-.0117106	-.0084376
c.age#c.age	.0001114	8.30e-06	13.43	0.000	.0000952	.0001277
1.female	.0357417	.0050851	7.03	0.000	.0257749	.0457085
1.nonwhite	-.0975841	.0065139	-14.98	0.000	-.1103515	-.0848168
1.married	.3080607	.0052834	58.31	0.000	.2977052	.3184162
educat						
1	.0678076	.0070752	9.58	0.000	.0539401	.0816751
2	.1473213	.0066123	22.28	0.000	.1343612	.1602814

```
ologit hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
```

hap	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9680552	.0026094	-12.04	0.000	.9629543	.9731831
c.age#c.age	1.000362	.0000268	13.50	0.000	1.00031	1.000415
1.female	1.121912	.0182835	7.06	0.000	1.086644	1.158326
1.nonwhite	.7349438	.0155689	-14.54	0.000	.7050541	.7661005
1.married	2.680555	.046998	56.24	0.000	2.590005	2.77427
educat						
1	1.231638	.0282539	9.08	0.000	1.177488	1.288278
2	1.578385	.033964	21.21	0.000	1.513201	1.646377

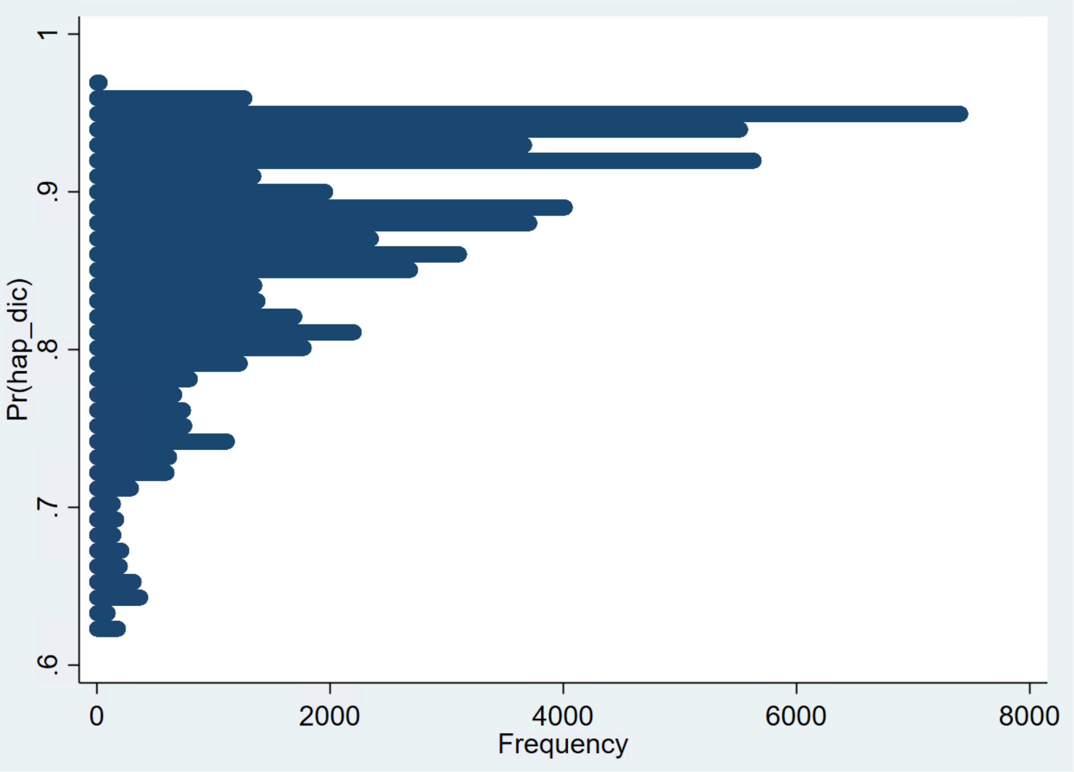
Distribution of predicted probabilities: logit

- What to expect

```
logit hap_dic c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
/*store predicted probabilities*/
predict prlogit if e(sample)==1 /*"e(sample) stata generated for "in model"*/
sum prlogit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
prlogit	59,725	.8725157	.0722409	.6214635	.967612

```
dotplot prlogit
```



- What are these?
- Interpretation?
- How can relate to AMEs?

- Can use to contextualize results

Distribution of predicted probabilities: ologit

- What to expect

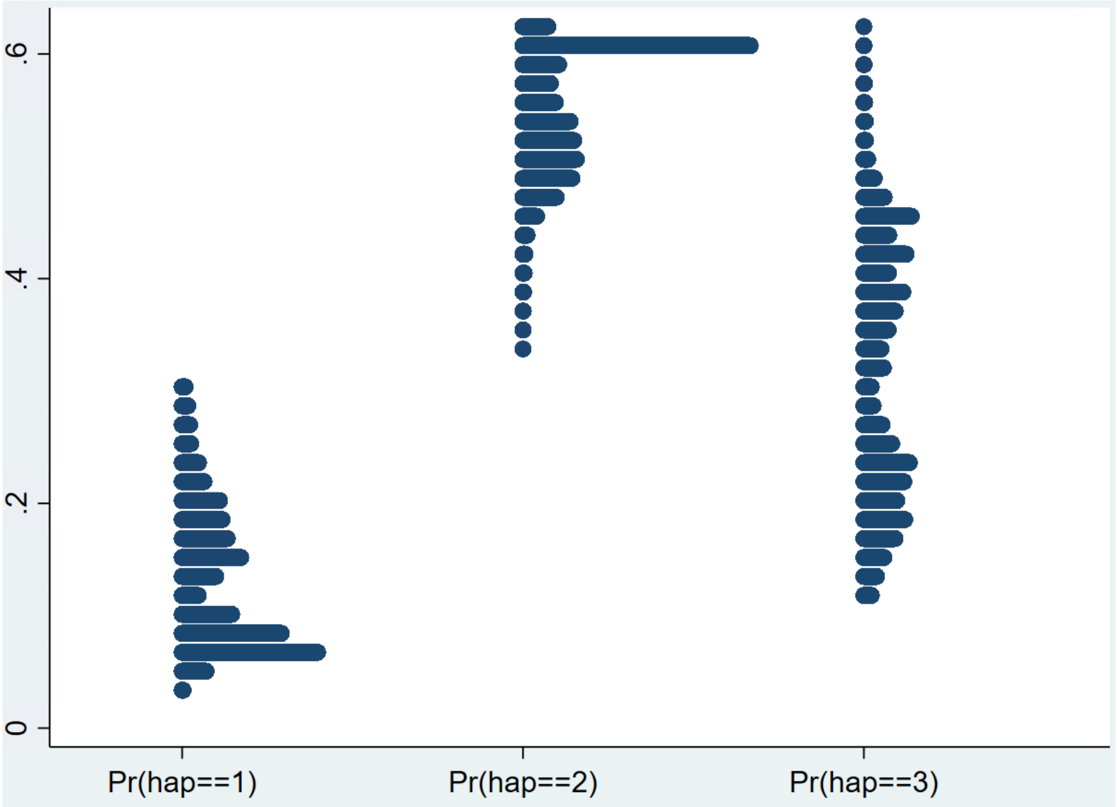
```
ologit hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
/*store predicted probabilities: create one var for each outcome category*/
predict prnot prpretty prvery if e(sample)==1
sum prnot prpretty prvery
```

Variable	Obs	Mean	Std. Dev.	Min	Max
prnot	59,725	.1279784	.0608348	.0329845	.3089663
prpretty	59,725	.5595987	.0540021	.343503	.6159315
prvery	59,725	.3124229	.1103339	.1121729	.6235124

- What are these?
- Interpretation?
- How can relate to AMEs?

- Can use to contextualize ideal types

```
dotplot prnot prpretty prvery
```



- How different from logit?
- Why?

Predicted probabilities: Marginal effects

- Marginal effect: Δ in the predicted probability given a Δ in X

- Average marginal effect (AME): the average of the marginal effect for all observations

- Marginal effect at the mean (MEM): all other Xs held at their means

- Marginal effect at representative values (MER): all other Xs held at substantively meaningful values

- Why/how is this an advantageous approach compared to odds ratios?

- What are the disadvantages of each approach, respectively (i.e., AME, MEM, MER)?

Predicted probabilities: Marginal effects

- Marginal effect: Δ in the predicted probability given a Δ in X
  - holding all other Xs constant

```
logit hap_dic c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
/*AME*/ mchange, decimals(4)
/*MEM*/ mchange, atmeans decimals(4)
/*MER*/ mchange, at(married=0 age=40 female=1 nonwhite=0 educat=1) decimals(4)
```

Logit	AME	MEM	MER
Age	-0.0004	-0.0004	-0.0013
Female	0.01	0.01	0.01
Nonwhite	-0.05	-0.05	-0.08
Married	0.11	0.12	0.11
Edu 1 vs 0	0.05	0.06	0.07
Edu 2 vs 0	0.09	0.09	0.12
Edu 2 vs 1	0.04	0.04	0.05

- Interpretation?

Predicted probabilities: Marginal effects

- Marginal effect: Δ in the predicted probability given a Δ in X
  - holding all other Xs constant

```
ologit hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
/*AME*/ mchange, decimals(4)
/*MEM*/ mchange, atmeans decimals(4)
/*MER*/ mchange, at(married=0 age=40 female=1 nonwhite=0 educat=1) decimals(4)
```

Ologit	AME 1	AME 2	AME 3	MEM 1	MEM 2	MEM 3	MER 1	MER 2	MER 3
Age	-0.0001	-0.0002	0.0003	-0.0001	-0.0002	0.0003	0.0005	0.0000	-0.0005
Female	-0.01	-0.1	0.02	-0.01	-0.1	0.02	-0.02	0.001	0.02
Nonwhite	0.04	0.02	-0.06	0.04	0.02	-0.06	0.052	-0.01	-0.04
Married	-0.11	-0.10	0.20	-0.11	-0.08	0.19	-0.11	-0.09	0.20
Edu 1 vs 0	-0.03	-0.01	0.04	-0.03	-0.01	0.04	-0.03	0.004	0.03
Edu 2 vs 0	-0.05	-0.04	0.09	-0.05	-0.04	0.09	-0.07	-0.002	0.07
Edu 2 vs 1	-0.03	-0.03	0.05	-0.03	-0.02	0.05	-0.04	-0.006	0.04

- Interpretation?

Predicted probabilities: Subgroups

- Not everyone is average (AME)
- Sometimes means don't make sense (MEM)
- Meaningful values for "ideal types" can be overwhelming (MER)
- Consider setting covariate values at subgroup means
  - And computing predicted probabilities

```
gen _sel40W = age==40 & nonwhite==0
label var _sel40W "40yr whites"
gen _sel40N = age==40 & nonwhite==1
label var _sel40N "40yr non-whites"
```

```
ologit hap c.age##c.age i.female i.nonwhite i.married i.educat if nmiss==0, or
mtable if _sel40W==1, rowname(1 40yr whites) atmeans ci
mtable if _sel40N==1, rowname(2 40yr non-whites)atmeans ci below
```

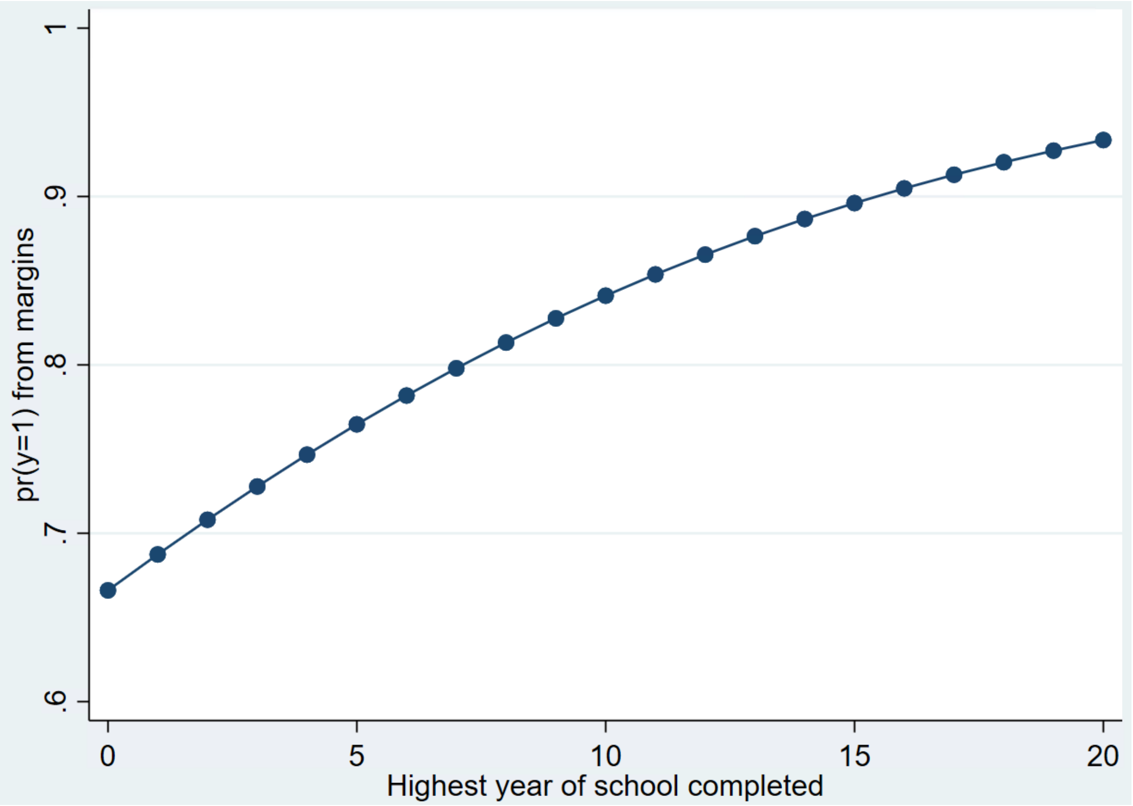
	age	1. female	nonwhite	1. married	1. educat	2. educat
Set 1	40	.523	0	.67	.297	.575
Current	40	.5	1	.422	.27	.496

	1	2	3
1 40yr whites_Pr(y)	0.103	0.567	0.330
1 40yr whites_ll	0.100	0.563	0.325
1 40yr whites_ul	0.106	0.572	0.335
2 40yr non-whites_Pr(y)	0.172	0.614	0.213
2 40yr non-whites_ll	0.166	0.610	0.207
2 40yr non-whites_ul	0.178	0.619	0.220

- Interpretation?

Predicted probabilities: graphing

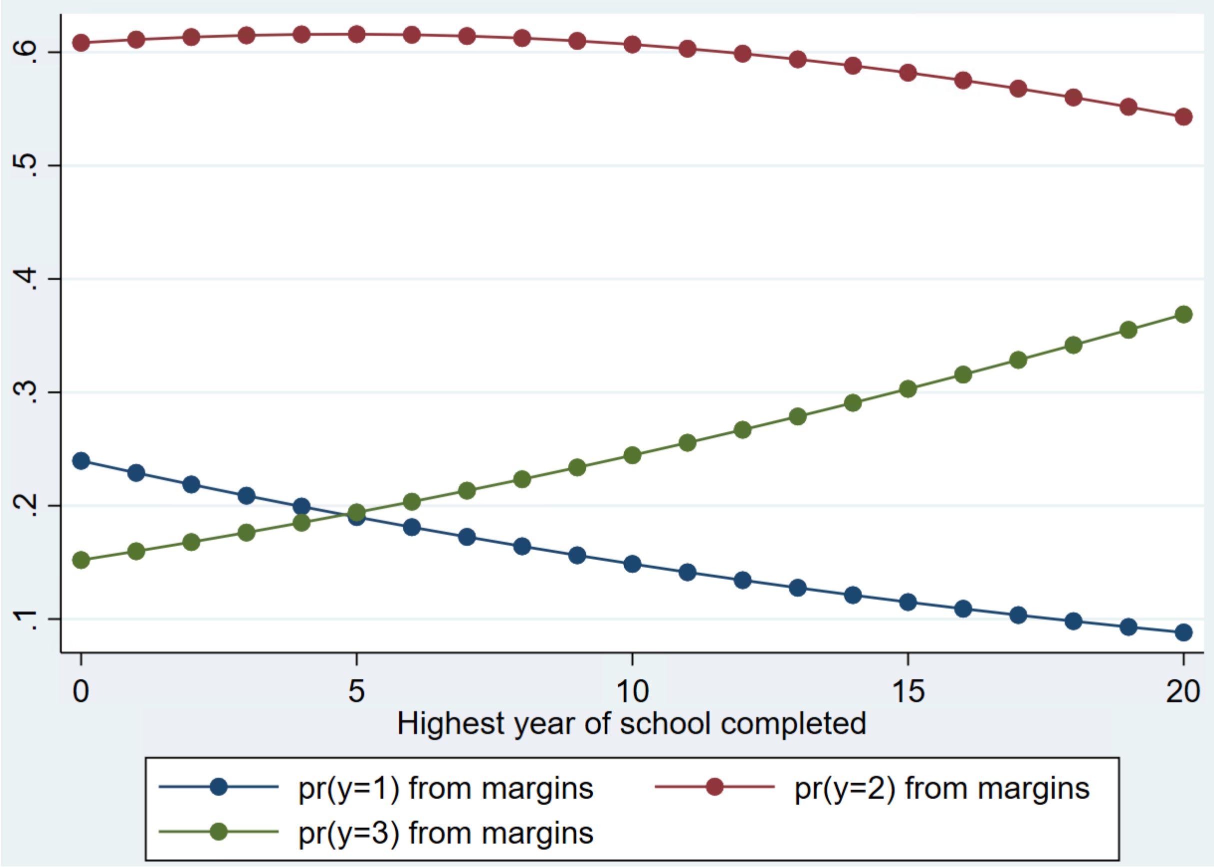
```
logit hap_dic c.age##c.age i.female i.nonwhite i.married c.educ if nmiss==0, or
mgen, at(educ=(0(1)20)) atmeans stub(CL0_)
graph twoway connected CL0_pr1 CL0_educ
```



- Interpretation?

Predicted probabilities: graphing

```
ologit hap c.age##c.age i.female i.nonwhite i.married c.educ if nmiss==0, or
mgen, at(educ=(0(1)20)) atmeans replace stub(CL1_)
graph twoway connected CL1_pr1 CL1_pr2 CL1_pr3 CL1_educ
```



- Interpretation?

Next class we will…

- Multinomial logistic intro
  - read Hoffmann CH5 & L&F CH8 before class