

Solutions for Session 8

29/11/2022

```
. do solution.do
. global basedir http://personalpages.manchester.ac.uk/staff/mark.lunt
. global datadir $basedir/stats/8_categorical/data
. use "$datadir/alligators.dta", clear

. label list
lake:
      1 Hancock
      2 Oklawaha
      3 Trafford
      4 George
gender:
      0 Male
      1 Female
size:
      0 <= 2.3m
      1 > 2.3m
food:
      1 Fish
      2 Invertebrate
      3 Reptile
      4 Bird
      5 Other

. gen invertebrate = food - 1 if food < 3
(64 missing values generated)
```


1.6 No: OR = 2.4, but p = 0.109

```
. mlogit food size, rrr
Iteration 0: log likelihood = -302.18146
Iteration 1: log likelihood = -294.795
Iteration 2: log likelihood = -294.60682
Iteration 3: log likelihood = -294.60668
Iteration 4: log likelihood = -294.60668
Multinomial logistic regression      Number of obs =      219
LR chi2(4) =      15.15
Prob > chi2 =      0.0044
Pseudo R2 =      0.0251
Log likelihood = -294.60668
```

food	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Fish	(base outcome)					
Invertebrate size	.3871605	.138164	-2.66	0.008	.1923654	.779211
_cons	.9183673	.1896165	-0.41	0.680	.6127281	1.376465
Reptile size	2.359259	1.262193	1.60	0.109	.8267738	6.732318
_cons	.122449	.0529618	-4.86	0.000	.0524556	.2858368
Bird size	1.742222	1.056353	0.92	0.360	.5308826	5.717532
_cons	.1020408	.0479058	-4.86	0.000	.0406588	.2560902
Other size	.7450292	.3091513	-0.71	0.478	.3303434	1.680277
_cons	.3877551	.1047942	-3.51	0.000	.2283037	.6585702

1.7 Fish is used as the comparison group

1.8 The odds ratios are exactly the same, the confidence intervals the same to at least 5 significant figures

```
. lincom [Reptile]size - [Invertebrate]size, eform
( 1) - [Invertebrate]size + [Reptile]size = 0
```

food	exp(b)	Std. Err.	z	P> z	[95% Conf. Interval]	
(1)	6.09375	3.491628	3.15	0.002	1.982252	18.73314

1.9 Yes, OR = 6.1, p = 0.002

```
. gen rep_inv = food == 3 if food == 3 | food == 2
(139 missing values generated)
```

```
. logistic rep_inv size, or
```

```
Logistic regression          Number of obs =      80
                             LR chi2(1)      =     10.87
                             Prob > chi2     =     0.0010
Log likelihood = -38.418556   Pseudo R2      =     0.1240
```

rep_inv	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
size	6.09375	3.491628	3.15	0.002	1.982252	18.73314
_cons	.1333333	.0579485	-4.64	0.000	.0568846	.3125233

1.10 OR is identical, ci the same to 5 s.f.

```
. tabulate food lake, co chi2
```

Key
frequency
column percentage

food	lake				Total
	Hancock	Oklawaha	Trafford	George	
Fish	30 54.55	18 37.50	13 24.53	33 52.38	94 42.92
Invertebrate	4 7.27	19 39.58	18 33.96	20 31.75	61 27.85
Reptile	3 5.45	7 14.58	8 15.09	1 1.59	19 8.68
Bird	5 9.09	1 2.08	4 7.55	3 4.76	13 5.94
Other	13 23.64	3 6.25	10 18.87	6 9.52	32 14.61
Total	55 100.00	48 100.00	53 100.00	63 100.00	219 100.00

Pearson chi2(12) = 37.7286 Pr = 0.000

1.11 There are big differences between the lakes in primary food choice (p = 0.000)

1.12 7.3%

1.13 The proportion was much higher in the other lakes

```

. mlogit food i.lake, rrr
Iteration 0:  log likelihood = -302.18146
Iteration 1:  log likelihood = -281.77642
Iteration 2:  log likelihood = -280.60116
Iteration 3:  log likelihood = -280.58389
Iteration 4:  log likelihood = -280.58384
Iteration 5:  log likelihood = -280.58384
Multinomial logistic regression          Number of obs   =       219
                                          LR chi2(12)    =       43.20
                                          Prob > chi2    =       0.0000
Log likelihood = -280.58384              Pseudo R2      =       0.0715

```

food	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Fish	(base outcome)					
Invertebrate lake						
Oklawaha	7.916667	4.953585	3.31	0.001	2.322373	26.98688
Trafford	10.38462	6.696362	3.63	0.000	2.93431	36.75148
George	4.545455	2.741012	2.51	0.012	1.394066	14.82078
_cons	.1333333	.0709721	-3.79	0.000	.0469734	.3784647
Reptile lake						
Oklawaha	3.888889	2.923346	1.81	0.071	.8911737	16.97027
Trafford	6.153846	4.640301	2.41	0.016	1.40378	26.97704
George	.3030303	.358162	-1.01	0.312	.0298828	3.072916
_cons	.1	.060553	-3.80	0.000	.0305191	.3276635
Bird lake						
Oklawaha	.3333333	.3784308	-0.97	0.333	.0360177	3.084902
Trafford	1.846154	1.381851	0.82	0.413	.4257368	8.005613
George	.5454545	.421439	-0.78	0.433	.1199742	2.479872
_cons	.1666667	.0805076	-3.71	0.000	.0646665	.4295544
Other lake						
Oklawaha	.3846154	.2717316	-1.35	0.176	.0963055	1.536039
Trafford	1.775148	.9512863	1.07	0.284	.6209878	5.074416
George	.4195804	.232565	-1.57	0.117	.1415831	1.243423
_cons	.4333333	.1438878	-2.52	0.012	.2260381	.8307351

1.14 This is highly significant, suggesting that there are big differences between the lakes in food preferences

1.15 7.92: Yes, the table suggested that preference for invertebrates was lowest in lake Hancock

. logistic invertibrate i.lake

Logistic regression
 Number of obs = 155
 LR chi2(3) = 19.48
 Prob > chi2 = 0.0002
 Pseudo R2 = 0.0938
 Log likelihood = -94.156736

invertibrate	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
lake						
Oklawaha	7.916666	4.953585	3.31	0.001	2.322373	26.98688
Trafford	10.38462	6.696362	3.63	0.000	2.93431	36.75148
George	4.545454	2.741012	2.51	0.012	1.394066	14.82078
_cons	.1333333	.0709721	-3.79	0.000	.0469734	.3784647

. clear

. use \$datadir/politics

. label list

party:
 1 Democrat
 2 Republican
 3 Independent

gender:
 0 Male
 1 Female

race:
 0 White
 1 Black

. mlogit party race, rrr

Iteration 0: log likelihood = -1088.7207
 Iteration 1: log likelihood = -1051.1104
 Iteration 2: log likelihood = -1049.6333
 Iteration 3: log likelihood = -1049.6121
 Iteration 4: log likelihood = -1049.6121

Multinomial logistic regression
 Number of obs = 1001
 LR chi2(2) = 78.22
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0359
 Log likelihood = -1049.6121

party	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Democrat	(base outcome)					
Republican						
race	.1017063	.034755	-6.69	0.000	.0520566	.1987101
_cons	1.003289	.0813108	0.04	0.968	.855936	1.176011
Independent						
race	.3258953	.0760263	-4.81	0.000	.2063032	.5148139
_cons	.8453947	.071637	-1.98	0.047	.7160287	.9981335

2.2 Blacks are less likely to be republicans rather than democrats when compare to whites
 OR = 0.10, p = 0.000
 2.3 Blacks are less likely to be independents rather than democrats when compare to whites
 OR = 0.33, p = 0.000

```
. mlogit party gender, rrr
Iteration 0: log likelihood = -1088.7207
Iteration 1: log likelihood = -1081.2678
Iteration 2: log likelihood = -1081.2533
Iteration 3: log likelihood = -1081.2533
Multinomial logistic regression      Number of obs   =      1001
                                      LR chi2(2)      =      14.93
                                      Prob > chi2     =      0.0006
Log likelihood = -1081.2533          Pseudo R2      =      0.0069
```

party	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Democrat	(base outcome)					
Republican						
gender	.5576923	.084847	-3.84	0.000	.413898	.7514428
_cons	1.045977	.1109013	0.42	0.672	.8497128	1.287574
Independent						
gender	.7961	.1239052	-1.47	0.143	.586795	1.080062
_cons	.7988506	.0908774	-1.97	0.048	.6391941	.9983857

2.5 0.56, 95% CI = 0.41, 0.75

```
. mlogit party i.race i.gender, rrr
```

```
Iteration 0: log likelihood = -1088.7207
Iteration 1: log likelihood = -1044.5431
Iteration 2: log likelihood = -1042.9153
Iteration 3: log likelihood = -1042.8912
Iteration 4: log likelihood = -1042.8912
```

Multinomial logistic regression

```
Number of obs = 1001
LR chi2(4) = 91.66
Prob > chi2 = 0.0000
Pseudo R2 = 0.0421
```

Log likelihood = -1042.8912

party	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Democrat	(base outcome)					
Republican						
race						
Black	.1024752	.0351279	-6.65	0.000	.0523398	.2006344
gender						
Female	.5639669	.0888365	-3.64	0.000	.4141634	.7679546
_cons	1.330487	.1498097	2.54	0.011	1.067007	1.659028
Independent						
race						
Black	.3268387	.076322	-4.79	0.000	.2068066	.5165384
gender						
Female	.8023691	.1269768	-1.39	0.164	.5883958	1.094155
_cons	.9514467	.1139876	-0.42	0.678	.752327	1.203268


```
. mlogit party i.race##i.gender, rrr
Iteration 0: log likelihood = -1088.7207
Iteration 1: log likelihood = -1044.6857
Iteration 2: log likelihood = -1042.8318
Iteration 3: log likelihood = -1042.7921
Iteration 4: log likelihood = -1042.7921
```

```
Multinomial logistic regression      Number of obs =      1001
LR chi2(6) =      91.86
Prob > chi2 =      0.0000
Pseudo R2 =      0.0422

Log likelihood = -1042.7921
```

party	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Democrat	(base outcome)					
Republican						
race						
Black	.1071429	.0483609	-4.95	0.000	.0442341	.2595189
gender						
Female	.5625	.0921255	-3.51	0.000	.4080504	.7754097
race#gender						
Black#Female	.8888959	.6191432	-0.17	0.866	.2269671	3.481279
_cons	1.333333	.1535221	2.50	0.012	1.063972	1.670888
Independent						
race						
Black	.2969629	.1039763	-3.47	0.001	.1495096	.5898415
gender						
Female	.7855704	.1336764	-1.42	0.156	.5627826	1.096553
race#gender						
Black#Female	1.1934	.560546	0.38	0.707	.4753052	2.996399
_cons	.9621212	.1195889	-0.31	0.756	.7540982	1.227529

```
. testparm i.race#i.gender
( 1) [Democrat]1o.race#1o.gender = 0
( 2) [Republican]1.race#1.gender = 0
( 3) [Independent]1.race#1.gender = 0
Constraint 1 dropped
      chi2( 2) = 0.20
      Prob > chi2 = 0.9061
```

*2.7 The interaction terms are not significant. So the effect of race is the same in men and women
Or, equivalently, the effect of gender is the same in blacks and whites*

```
. clear
. use $datadir/housing
```

. label list

housing:

- 1 Tower Block
- 2 Apartment
- 3 Atrium House
- 4 Terraced House

hml:

- 1 Low
- 2 Medium
- 3 High

. ologit satisfaction i.housing

Iteration 0: log likelihood = -1807.6794
 Iteration 1: log likelihood = -1779.8874
 Iteration 2: log likelihood = -1779.8691
 Iteration 3: log likelihood = -1779.8691

Ordered logistic regression

Number of obs = 1663
 LR chi2(3) = 55.62
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0154

Log likelihood = -1779.8691

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
housing						
Apartment	-.5214429	.1165401	-4.47	0.000	-.7498572	-.2930286
Atrium House	-.2813625	.1504761	-1.87	0.062	-.5762902	.0135653
Terraced House	-1.057463	.1476192	-7.16	0.000	-1.346791	-.7681346
/cut1	-1.126582	.0998234			-1.322233	-.9309321
/cut2	.0071965	.0957355			-.1804416	.1948345

3.2 Yes: the fit of the model is highly significant ($chi2 = 55.6, p = 0.0000$)

3.3 The most satisfied are the ones with the highest coefficient. In this case, this is group 1 (Tower Blocks)

. ologit satisfaction influence

Iteration 0: log likelihood = -1807.6794
 Iteration 1: log likelihood = -1755.9356
 Iteration 2: log likelihood = -1755.8646
 Iteration 3: log likelihood = -1755.8646

Ordered logistic regression

Number of obs = 1663
 LR chi2(1) = 103.63
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0287

Log likelihood = -1755.8646

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
influence	.6130794	.061265	10.01	0.000	.4930022	.7331566
/cut1	.4431937	.1209239			.2061872	.6802003
/cut2	1.604388	.1270715			1.355333	1.853444

3.4 Influence is a highly significant predictor of satisfaction
 I have fitted it as a continuous variable, assuming a trend in the effect.
 Fitting it as a categorical variable confirms that this is reasonable

```
. ologit satisfaction i.influence
Iteration 0:  log likelihood = -1807.6794
Iteration 1:  log likelihood = -1755.1883
Iteration 2:  log likelihood = -1755.106
Iteration 3:  log likelihood = -1755.106

Ordered logistic regression      Number of obs =      1663
                                LR chi2(2)         =     105.15
                                Prob > chi2          =      0.0000
                                Pseudo R2           =      0.0291

Log likelihood = -1755.106
```

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
influence						
Medium	.5093184	.1040864	4.89	0.000	.3053128	.7133239
High	1.253065	.1249063	10.03	0.000	1.008253	1.497877
/cut1	-.2059269	.0767168			-.356289	-.0555647
/cut2	.9558175	.0804169			.7982033	1.113432

The effect of high influence is about twice the effect of medium influence, so are trend model is appropriate

```
. ologit satisfaction contact
Iteration 0:  log likelihood = -1807.6794
Iteration 1:  log likelihood = -1806.7024
Iteration 2:  log likelihood = -1806.7024

Ordered logistic regression      Number of obs =      1663
                                LR chi2(1)         =       1.95
                                Prob > chi2          =     0.1622
                                Pseudo R2           =     0.0005

Log likelihood = -1806.7024
```

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
contact	.1280512	.0916176	1.40	0.162	-.051516	.3076184
/cut1	-.5857037	.0736012			-.7299595	-.4414479
/cut2	.5181326	.0732786			.3745091	.661756

3.4 Contact is not a significant predictor of satisfaction

```

. ologit satisfaction influence i.housing
Iteration 0: log likelihood = -1807.6794
Iteration 1: log likelihood = -1730.027
Iteration 2: log likelihood = -1729.8401
Iteration 3: log likelihood = -1729.84
Ordered logistic regression
Log likelihood = -1729.84
Number of obs = 1663
LR chi2(4) = 155.68
Prob > chi2 = 0.0000
Pseudo R2 = 0.0431

```

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
influence	.6082232	.0618167	9.84	0.000	.4870646	.7293817
housing						
Apartment	-.5717304	.118653	-4.82	0.000	-.804286	-.3391749
Atrium House	-.2857831	.1533433	-1.86	0.062	-.5863305	.0147643
Terraced House	-1.016881	.1497045	-6.79	0.000	-1.310297	-.7234658
/cut1	-.0512936	.1473787			-.3401506	.2375633
/cut2	1.140398	.1501475			.8461145	1.434682

3.5 Influence is still significant after adjusting for housing type (z = 9.84, p= 0.000)

```

. testparm i.housing
( 1) [satisfaction]2.housing = 0
( 2) [satisfaction]3.housing = 0
( 3) [satisfaction]4.housing = 0
      chi2( 3) = 51.01
      Prob > chi2 = 0.0000

```

3.5 Housing is still significant after adjusting for influence (chi2 = 51, p= 0.0000)

```
. ologit satisfaction i.housing#c.influence
```

```
Iteration 0: log likelihood = -1807.6794
Iteration 1: log likelihood = -1724.1018
Iteration 2: log likelihood = -1723.9011
Iteration 3: log likelihood = -1723.901
```

```
Ordered logistic regression
```

```
Number of obs = 1663
LR chi2(7) = 167.56
Prob > chi2 = 0.0000
Pseudo R2 = 0.0463
```

```
Log likelihood = -1723.901
```

satisfaction	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
housing						
Apartment	-1.40195	.3145598	-4.46	0.000	-2.018476	-.7854243
Atrium House	-.3543242	.3938297	-0.90	0.368	-1.126216	.4175677
Terraced House	-1.824184	.3956123	-4.61	0.000	-2.59957	-1.048799
influence	.3253134	.1288222	2.53	0.012	.0728266	.5778002
housing#c.influence						
Apartment	.4492471	.1573122	2.86	0.004	.140921	.7575733
Atrium House	.0376429	.1986514	0.19	0.850	-.3517067	.4269925
Terraced House	.4481545	.2041759	2.19	0.028	.0479772	.8483318
/cut1	-.5706535	.2555718			-1.071565	-.0697419
/cut2	.6288746	.2554054			.1282893	1.12946

```
. testparm i.housing#c.influence
```

```
( 1) [satisfaction]2.housing#c.influence = 0
( 2) [satisfaction]3.housing#c.influence = 0
( 3) [satisfaction]4.housing#c.influence = 0
```

```
chi2( 3) = 11.92
Prob > chi2 = 0.0077
```

3.6 Yes, the three interaction terms are significant. Influence is of greater importance in housing types 2 and 4 than in housing types 1 and 3
end of do-file