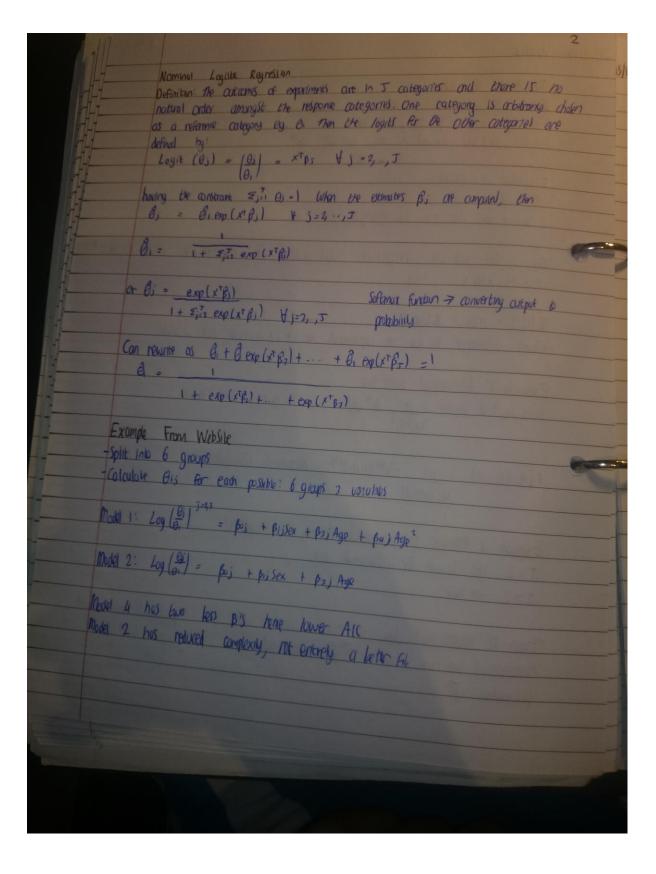
,	Could be types.	12/16/16 DM
08/03/16	ALSM2	N. American
	Multinamial Oscarlation  Extension to Grammal Dysolutar  Grammal is a joint probability distribution  P (4, , 4, 10, , 0, n) - n! 6, 8, 6, 6, 6, 6, 7	
0	S: number of cutescres.  For except J=3 yes to must   Y1: # yp) Y2: # no Y2: # musts  Oi,, Or are the respective probabilities of the cutes  N = Y1 + Y2 + + YJ	w be 0,00 + er = 1
	E[y,] = no. E[y,] = no Main difference so for > represent variohe in a vector, h	pt a side
	Exercise 1: Multinomial with 5=2  P(4, 4, 10, 02/h) = n! 0, 4, 6, 4, 12  9, 14, 2	west and along
•	$9(9(19(n)) = n! + \theta_1^{9}(1-9)^{n-9} + binomial Distribution  9((ny_1)!)$	
	Exercise 2: Member of experential family of distributors?  T = 2 yes, Shown for bironnal dist.  T = 2 NO	
0	tion if multinanial dist is not a member of the distributions, we can arrival By collected over parameters B	exponential family of or groups via a set of

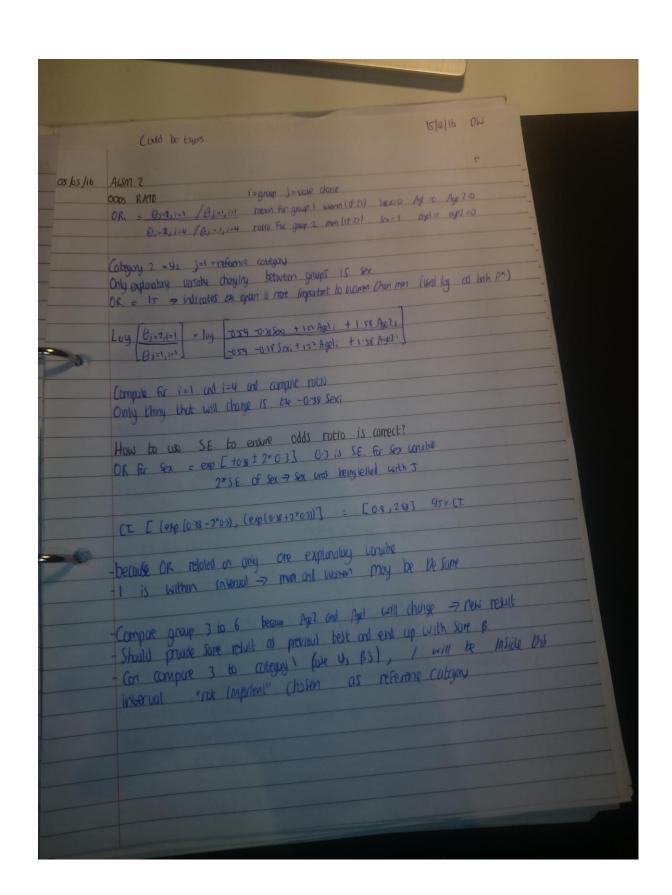


2 Codd be types.	12/10/14 OM
08/03/16 ALSM 2  R output: (More) 17  y 25ex +9xl +aye  interant 8x agel ax 2  42 -059 -039 113 159  93 -104 -081 1149 292	3
What are the $0$ 's for model 1?  Use softmax fuxton, $\beta_i = \exp\left(\frac{x^{\gamma}\beta_i}{x^{\gamma}}\right)$ $1 + \sum_{j=1}^{\infty} \exp(x^{\gamma}\beta_j)$ $1 + \sum_{j=1}^{\infty} \exp(x^{\gamma}\beta_j)$ $1 + \sum_{j=1}^{\infty} \exp(x^{\gamma}\beta_j)$ $1 + \sum_{j=1}^{\infty} \exp(x^{\gamma}\beta_j)$	
$\frac{3}{5}$ $\frac{4}{5}$ $\frac{6}{5}$ $\frac{6}{5}$ $\frac{1}{5}$ $\frac{1}$	
(3) - 3j -	al ampl
For $\hat{\theta}_2$ (the overall $y_2$ value, NOT per individue $\hat{\theta}_{21} = \exp(-0.59 - 0.94(50.1) + 1.17(490.1) + 1.59(690.2)$ $1 + \exp(-0.59 - 0.345.00 + 1.13(490.1) + 1.59(490.21) + exp(-1.04)$	anin's Calvary's
Same for B3 & is 1-82-B5	

ODDS RATIO women important/women not important (18-23) B21 / On D =4/614 men inspetut from not important 18-23 ald ratio assume we of softmax function -Papertan of woman of important ove not important is bigger than that of men larto vote important over not important for age group 18-23. Multinomial Distribution histered of value of response, hove a vector of responses (y) Obene a number of groups with vector response and # page in each group { (yi, xi, ni) } i = >,..., N (number of groups)

| 15 # people today in goap i.
| Explanatory limite associated with goap i. Johnson, II df. #B(T-1) = O.F. if intercept, sox, age = 3x2 = 6 D.F. in example DF of solvoid model 2 of for each grap (3 upon pergrap) 6 groups = 12 OF. At Subratal resta When All distral by 2 between models, could be form on extra variouse (2m) in All famula, m is mil, hene extra 2 Company MI and mz M. 8 property, M2-6

Gilleren between log litelihoods -2[kg/km - Laylons] \$2 Are the moders different? 2 ~ x? 2 will be in the 95% at of the square we can say M 25m2, basically quivalent.



	1
2	
19 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
$A = \pi_{i}^{16} \qquad \theta_{1i} \qquad \theta_{1i} \qquad \theta_{2i} \qquad \theta_$	4.
Florida Company	
A(O, Vo, O, O, O, O) de = 6x4	
L ( fp. 3 3=2 15)	
E ALC-contertorn to seject bust makel. Can use to drup parameted or not	
F. P. X = 0.17 - 0.4680x - 1345128 - 1-786, + 0.4162 + 1.1623	100
β\$ X = -342 -0.638x +0.765128 +1.154 +2.7312 + 7.066,	
FEX = -243 -0618xx + 0735/120 +0754 -05762 +12463	4
1 - 143 -0215px -0245pe +0774 +0.0264, +1.5644	
G. We take of fitted vale libre g	
Hencoch rs Ohlawha COD RATIO	
13 UN (OWNA COD RATIO  13 000/03 = 0006 3109	10
j=2 01/612 [=6 0-023/649 3:03	
1=3 0074/671 1=7 001/621 309 1=4 071/672 1=3 0438/64	18.18
$OR = \exp\left(\beta_{j=0} \alpha \beta_{j=0} \alpha\right) \approx 113$	
H Curiulat?	3
$SE, \beta_{3}=6 = 0.79 \qquad \beta_{3}=9 = 12$	
Le: 05/4 45=4 = 12	O PULL
$\frac{ ak ^2}{ ak ^2} \frac{SE(p_1^{24} p_2^{24})^2}{ ak ^2} = \frac{SE(p_1^{24})^2}{ bk ^2} + \frac{SE(p_1^{24})^2}{ ak ^2} = \frac{1}{ ak ^2} \frac{ ak ^2}{ ak ^2} = \frac{1}{ ak ^2}  ak $	
maxicular assume.	1
S.E = 10.79° +12° =145	1
113 ± 145 ×2 = 95%. CI	
D within the interval > mt antival	
Cy (0) =1, there is a characteristic	100
Cont sy for five bot "ally or) in Our in lake "	1
that alligar) in than in lap	1
	-
	-

## 4. Alligators

The table 7 comes from a study of the primary food choices of alligators in four Florida lakes. Researchers classified the stomach contents of 219 captured alligators into five categories. Fish (the most common primary food choice), Invertebrate (snails, insects, crayfish, etc.). Reptile (turties, alligators), Bird, and Other (amphibians, plants, household pets, stones, and other debris).

ī	Lake	Sex	Size	Fish	Inver.	Rept.	Bird	Other
-				$y_1$	y2	У3	<i>y</i> 4	<i>y</i> 5
1	Hancock	M	small	7	1	0	0	5
2			large	4	0	0	1	2
3		F	small	16	3	2	2	3
4			large	3	0	1	2	3
5	Oklawaha	M	small	2	2	0	0	1
6			large	13	7	6	0	0
7		F	small	3	9	1	0	2
8			large	0	1	0	1	0
9	Trafford	M	small	3	7	1	0	1
10			large	8	6	6	3	5
11		F	small	2	4	- 1	1	4
12			large	0	1	0	0	0
13	George	M	small	13	10	0	2	2
14			large	9	0	0	1	2
15		F	small	3	9	1	0	1
16			large	В	1	0	0	1

Table 7: Alligators primary Food Choice.

The expert decides to use the multinomial distribution to model the responses  $\{y_1,\cdots,y_5\}$ .

- (a) Explain why this distribution is suited for this problem.
- (b) The expert defines the following explanatory variables:

$$sex = \begin{cases} 0 & female \\ 1 & male \end{cases} \quad size = \begin{cases} 0 & small \\ 1 & large \end{cases} \quad Lake_1 = \begin{cases} 1 & If Hancock \\ 0 & otherwise \end{cases}$$

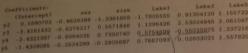
$$Lake_2 = \begin{cases} 1 & If Oklawaha \\ 0 & otherwise \end{cases} \quad Lake_3 = \begin{cases} 1 & If Trafford \\ 0 & otherwise \end{cases}$$

Explain why only 3 indicator variables are defined to encode the information about the 4 possible lakes where the data has been collected?

- (c) Assuming the reference category is  $\theta_1$  (proportion for *fish*), write down the linear model linking the proportions  $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$  with the explanatory variables.
- (d) Explain what function is used and maximised to compute the parameters of the models (noted  $\beta$ s in the course). Write down explicitly this function with the notation you have introduced in question 4c.
- (e) The model is fitted using R and the output in R is:

Call

multinom(formula = y.mat " sex + size + Lake1 + Lake2 + Lake3)



Explain the meaning of the AIC and its use.

- (f) Rewrite the model find in question 4c using the numerical values found with R.
- (g) The fitted values are also computed automatically with R and these are reported in table 8.

	Figh	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\tilde{ heta}_5$
-		0.07545645	0.032585176	0.051157366	0.24017062
1		0.07543843	0.059063749	0.110228530	0.18648582
2	0.5070764	0.10120786	0.051529843	0.079201241	0.26098463
4	0.5070764	0.02705796	0.091497934	0.167174242	0.19851457
-5	0.3137353	(0.56356125	0.066792314	0.008384288	0.05784657
6	0.4825231	0.23557649	0.185433818	0.027670279	0.06879629
7	0.2146236	0.63333355	0.088499072	0.010875857	0.05266792
_8	0.3591723	0.27859182	0.258550999	0.037770774	0.06591415
9	0.2088132	0.49437446	0.078160321	0.034470791	0.18418127
10	0.3050677	0.18984756	0.199345737	0.104509741	0.20122923
11	0.1449154	0.54508513	0.101605098	0.043869783	0.16452463
12	0.2132216	0.21081022	0.260984348	0.133952044	0.18103180
13	0.5008607	0.37332877	0.008780762	0.023993546	0.09303619
14	0.6826618	0.13374903	0.020893106	0.067865670	0.09483043
15	0.3930845	0.46549198	0.012908441	0.034531949	0.09398311
16	0.5781329	0.17995524	0.033143452	0.105397407	0.10337101

## Table 8: Estimated fitted values.

The expert says: alligators in Lake Oklawaha are less likely to choose birds over fish than their colleagues in Lake Hancock are. Explain mathematically where this conclusion comes from

(h) Assuming the  $\beta$ s independent, how confident are you in this conclusion by the expert given the standard errors of the parameters computed by R:

 Std. Errors:

 (Intercept)
 sex
 size
 Lake1
 Lake2
 Lake3

 y2
 0.3787475
 0.3955162
 0.4111827
 0.6232075
 0.4761068
 0.4927795

 y3
 1.0851582
 0.6852750
 0.6466092
 1.1928075
 1.1221413
 1.1297557

 y4
 0.7706720
 0.6888385
 0.6522657
 0.7952303
 1.2098680
 0.8661052

 y5
 0.5381162
 0.4663546
 0.4599317
 0.5685673
 0.7777958
 0.6256868

(25 marks)