Nominal logistic regression (Ch. 8):

8.3.1 Example: Car preferences

Table 8.1 *Importance of air conditioning and power steering in cars (row percent in brackets*)*

		Response			
Sex	Age	No or little importance	Important	Very important	Total
Women	18-23	26 (58%)	12 (27%)	7 (16%)	45
	24-40	9 (20%)	21 (47%)	15 (33%)	45
	> 40	5 (8%)	14 (23%)	41 (68%)	60
Men	18-30	40 (62%)	17 (26%)	8 (12%)	65
	24-40	17 (39%)	15 (34%)	12 (27%)	44
	> 40	8 (20%)	15 (37%)	18 (44%)	41
Total		105	94	101	300

```
no/little important very_important sex age
    26    12     7     1    1
     9    21    15     1    2
     5    14    41     1    3
    40    17     8     2    1
    17    15    12     2    2
     8    15    18     2    3
```

Table8.1.txt


```
> library(VGAM)
> data81 <- read.table(file="Table8.1.txt",header=T)
> data81 <- data.frame(data81)
> sapply(data81,class)
     no.little      important very_important            sex            age
     "integer"      "integer"      "integer"      "integer"      "integer"
> data81$sex <- factor(sex,labels=c("female","male"))
> data81$age <- factor(age,labels=c("18-23","24-40","41+"))
> sapply(data81,class)
     no.little      important very_important            sex            age
     "integer"      "integer"      "integer"       "factor"       "factor"
> data81
  no.little important very_important    sex   age
1        26        12              7 female 18-23
2         9        21             15 female 24-40
3         5        14             41 female   41+
4        40        17              8   male 18-23
5        17        15             12   male 24-40
6         8        15             18   male   41+
> m1 <- vglm(cbind(data81[,2],data81[,3],data81[,1]) ~ sex+age, multinomial, data81)
#
# NOTE: Referece category in R is the Last one, While Dobson uses the first
#
> coef(m1)
(Intercept):1 (Intercept):2     sexmale:1     sexmale:2    age24-40:1
   -0.5907988    -1.0390757    -0.3881281    -0.8130175     1.1282658
   age24-40:2      age41+:1      age41+:2
    1.4781069     1.5877072     2.9167514
```

$$x_1 = \begin{cases} 1 & \text{for men} \\ 0 & \text{for women} \end{cases}, \quad x_2 = \begin{cases} 1 & \text{for age 24-40 years} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{and } x_3 = \begin{cases} 1 & \text{for age} > 40 \text{ years} \\ 0 & \text{otherwise} \end{cases}.$$

Model (Dobson (8.10)):

$$\log(\frac{\pi_2}{\pi_1}) = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3$$

$$\log(\frac{\pi_3}{\pi_1}) = \beta_{03} + \beta_{13}x_1 + \beta_{23}x_2 + \beta_{33}x_3$$

In R-output:

(Intercept):1	sexmale:1	age24-40:1	age41+:1
β_{02}	β_{12}	β_{22}	β_{32}

(Intercept):2	sexmale: 2	age24-40:2	age41+:2
β_{03}	β_{13}	β_{23}	β_{33}


```
> summary(m1)
Call:
vglm(formula = cbind(data81[, 2], data81[, 3], data81[, 1]) ~
    sex + age, family = multinomial, data = data81)
Pearson Residuals:
  log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
1           -0.43958           -0.59747
2            0.89526           -0.13747
3           -0.25745            0.68642
4            0.35593            0.59464
5           -0.86024            0.27183
6            0.35191           -0.77493
Coefficients:
                 Value Std. Error t value
(Intercept):1 -0.59080    0.28398 -2.0804
(Intercept):2 -1.03908    0.33050 -3.1440
sexmale:1     -0.38813    0.30051 -1.2916
sexmale:2     -0.81302    0.32104 -2.5325
age24-40:1     1.12827    0.34165  3.3024
age24-40:2     1.47811    0.40092  3.6868
age41+:1       1.58771    0.40290  3.9407
age41+:2       2.91675    0.42292  6.8967
Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Dispersion Parameter for multinomial family:   1
Residual Deviance: 3.93871 on 4 degrees of freedom
Log-likelihood: -290.3511 on 4 degrees of freedom
Number of Iterations: 3
```

Minimal model

```
> m0 <- vglm(cbind(data81[,2],data81[,3],data81[,1]) ~ 1, multinomial, data81)
> summary(m0)
Call:
vglm(formula = cbind(data81[, 2], data81[, 3], data81[, 1]) ~
    1, family = multinomial, data = data81)
Pearson Residuals:
  log(mu[,1]/mu[,3]) log(mu[,2]/mu[,3])
1           -1.48156           -3.02865
2            2.43838            0.57029
3            0.14808            5.90016
4           -2.03766           -4.27693
5            0.17933           -0.88043
6            1.19833            1.73517
Coefficients:
                  Value Std. Error  t value
(Intercept):1 -0.11067    0.14199 -0.77937
(Intercept):2 -0.03884    0.13937 -0.27868
Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Dispersion Parameter for multinomial family:   1
Residual Deviance: 81.78056 on 10 degrees of freedom
Log-likelihood: -329.272 on 10 degrees of freedom
Number of Iterations: 4
```


Table 8.2 *Results of fitting the nominal logistic regression model (8.10) to the d in Table 8.1.*

Parameter β	Estimate b (std. error)	Odds ratio, $OR = e^b$ (95% confidence interval)	
$\log(\pi_2/\pi_1)$: important vs. no/little importance			
β_{02}: constant	-0.591 (0.284)		
β_{12}: men	-0.388 (0.301)	0.68	(0.38, 1.22)
β_{22}: 24-40	1.128 (0.342)	3.09	(1.58, 6.04)
β_{32}: >40	1.588 (0.403)	4.89	(2.22, 10.78)
$\log(\pi_3/\pi_1)$: very important vs. no/little importance			
β_{03}: constant	-1.039 (0.331)		
β_{13}: men	-0.813 (0.321)	0.44	(0.24, 0.83)
β_{23}: 24-40	1.478 (0.401)	4.38	(2.00, 9.62)
β_{33}: > 40	2.917 (0.423)	18.48	(8.07, 42.34)

Table 8.3 *Results from fitting the nominal logistic regression model (8.10) to the data in Table 8.1.*

Sex	Age	Importance Rating*	Obs. freq.	Estimated probability	Fitted value	Pearson residual
Women	18-23	1	26	0.524	23.59	0.496
		2	12	0.290	13.07	-0.295
		3	7	0.186	8.35	-0.466
	24-40	1	9	0.234	10.56	-0.479
		2	21	0.402	18.07	0.690
		3	15	0.364	16.37	-0.340
	> 40	1	5	0.098	5.85	-0.353
		2	14	0.264	15.87	-0.468
		3	41	0.638	38.28	0.440
Men	18-23	1	40	0.652	42.41	-0.370
		2	17	0.245	15.93	0.267
		3	8	0.102	6.65	0.522
	24-40	1	17	0.351	15.44	0.396
		2	15	0.408	17.93	-0.692
		3	12	0.241	10.63	0.422
	> 40	1	8	0.174	7.15	0.320
		2	15	0.320	13.13	0.515
		3	18	0.505	20.72	-0.600
Total			300		300	
Sum of squares						3.931

An alternative model can be fitted with age group as covariate, that is

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2; \quad j = 2, 3, \tag{8.11}$$

where

$$x_1 = \begin{cases} 1 & \text{for men} \\ 0 & \text{for women} \end{cases} \quad \text{and} \quad x_2 = \begin{cases} 0 & \text{for age group 18-23} \\ 1 & \text{for age group 24-40} \\ 2 & \text{for age group} > 40 \end{cases}$$

This model fits the data almost as well as (8.10) but with two fewer parameters. The maximum value of the log likelihood function is -291.05 so the difference in deviance from model (8.10) is

$$\triangle D = 2 \times (-290.35 + 291.05) = 1.4$$

which is not significant compared with the distribution $\chi^2(2)$. So on the grounds of parsimony model (8.11) is preferable.