Slide 1

Reference: Agresti, Sections 6.1, 6.3.2

Multinomial responses

- We now turn attention to regression models for the analysis of categorical response data with more than two categories.
- Logistic regression can be extended to handle responses that are *polytomous*, i.e., with $c > 2$ categories.
- We first consider models for **nominal** response data.
- Next, we will look at models for **ordinal** response data.

Slide 2

Multinomial distribution

- Suppose Y_i takes several discrete values, $j = 1, 2, \ldots, c$

$$\pi_{ij} = p(Y_i = j).$$

- For individual data ($n_i = 1$), let Y_{ij} denote an indicator variable that takes value 1 if the i-th response falls in the j-th category and 0 otherwise ($\sum_j y_{ij} = 1$).
- For grouped data, Y_{ij} denote the number of responses from the i-th group that fall in the j-th category ($\sum_j y_{ij} = n_i$).
- The probability distribution of the counts Y_{ij} given the total n_i is given by the **multinomial distribution**

$$p(Y_{i1} = y_{i1}, \ldots, Y_{ic} = y_{ic}) = \binom{n_i}{y_{i1}, \ldots, y_{ic}} \pi_{i1}^{y_{i1}} \ldots \pi_{ic}^{y_{ic}}.$$

Slide 3

- The multinomial distribution is a member of the multivariate exponential family.
- Since $y_{ic} = 1 - (y_{i1} + \ldots + y_{i,c-1})$ and $\pi_{ic} = 1 - \sum_{j=1}^{c-1} \pi_{ij}$

$$
\begin{aligned}
f(\boldsymbol{y}_i) &= \exp\left\{ \sum_{j=1}^{c-1} y_{ij} \log \pi_{ij} + \left(1 - \sum_{j=1}^{c-1} y_{ij}\right) \log \pi_{ic} + \log \binom{1}{y_{i1}, \ldots, y_{ic}} \right\} \\
&= \exp\left\{ \sum_{j=1}^{c-1} y_{ij} \log \frac{\pi_{ij}}{\pi_{ic}} + \log \pi_{ic} + \log \binom{1}{y_{i1}, \ldots, y_{ic}} \right\} \\
&= \exp\left\{ \boldsymbol{y}_i^T \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) + c(\boldsymbol{y}_i, \phi) \right\}
\end{aligned}
$$

where

$$
\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i,c-1})^T, \ \boldsymbol{\theta}_i = \left(\log \frac{\pi_{i1}}{\pi_{ic}}, \ldots, \log \frac{\pi_{i,c-1}}{\pi_{ic}} \right).
$$

Slide 4

Multinomial logit model

- Thus, the corresponding GLM with canonical link is given by

$$
\log\left(\frac{\pi_{ij}}{\pi_{ic}}\right) = \boldsymbol{x}_i' \boldsymbol{\beta}_j = \sum_{k=1}^{p} \beta_{jk} x_{ik}, \quad j = 1, \ldots, c-1.
$$

- This is called the **baseline category model** or **multinomial logit model**; it calculates the log-odds for all other categories relative to the baseline.
- If $\boldsymbol{x}_i$ has length p then $\boldsymbol{\beta}_j$ is a p-vector of regression coefficients.
- β_{jk} is the log-odds ratio for response j versus the baseline for a one unit increase in covariate x_{ik}.

Slide 5

- The comparison of two pairs of response categories, j and j', is given by:

$$\log \frac{\pi_{ij}}{\pi_{ij'}} = \log \frac{\pi_{ij}}{\pi_{ic}} - \log \frac{\pi_{ij'}}{\pi_{ic}} = \boldsymbol{x}_i'(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j'})$$

$\beta_{jk} - \beta_{j'k}$ is the log-odds ratio for response j versus j' for a one unit increase in covariate x_{ik}.

- For the reference group,

$$\pi_{ic} = \frac{1}{1 + \sum_{h=1}^{c-1} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_h)}.$$

- For the other groups,

$$\pi_{ij} = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{h=1}^{c-1} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_h)}$$

Slide 6

- ML estimates for the parameters can be found by maximizing the multinomial likelihood with the probabilities π_{ij} viewed as functions of the $\boldsymbol{\beta}_j$'s.
- The log-likelihood function is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{c-1} y_{ij} \left(\boldsymbol{x}_i'\boldsymbol{\beta}_j \right) - \log \left[1 + \sum_{j=1}^{c-1} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \right] \right\}$$

- The score equation is

$$\frac{\partial \log L(\boldsymbol{\beta})}{\partial \beta_{jk}} = \sum_{i=1}^{N} x_{ik} y_{ij} - \sum_{i=1}^{N} \left[\frac{x_{ik} \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{h=1}^{c-1} \exp(\boldsymbol{x}_i \boldsymbol{\beta}_h)} \right] = \sum_{i=1}^{N} x_{ik}(y_{ij} - \pi_{ij}).$$

Slide 7

- The Hessian matrix consists of $(c-1)^2$ blocks of size $p \times p$ with elements

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{jk'}} = -\sum_{i=1}^{N} x_{ik} x_{ik'} \pi_{ij}(1-\pi_{ij}), \qquad \text{within same } j$$

$$\frac{\partial^2 \log L(\boldsymbol{\beta})}{\partial \beta_{jk} \partial \beta_{j'k'}} = \sum_{i=1}^{N} x_{ik} x_{ik'} \pi_{ij} \pi_{ij'}, \qquad \text{for } j \neq j'$$

- Since the baseline-category logits are the canonical links, the Newton-Raphson and the Fisher scoring algorithms are equivalent.

Slide 8

Goodness of fit

- The deviance statistic can be used to test the fit of a model versus a saturated model

$$G^2 = 2 \sum_{i=1}^{N} \sum_{j=1}^{c} n_i y_{ij} \log \frac{n_i y_{ij}}{n_i \hat{\pi}_{ij}}.$$

- The corresponding Pearson statistic is

$$\mathcal{X}^2 = \sum_{i=1}^{N} \sum_{j=1}^{c} \frac{(n_i y_{ij} - n_i \hat{\pi}_{ij})^2}{n_i \hat{\pi}_{ij}}.$$

- For grouped data, both are asymptotically distributed as χ^2_{df} with

$$df = (N-p)(c-1)$$

since the saturated model has $N(c-1)$ free parameters and the current model has $p(c-1)$.

Slide 9

Example

- Consider Table 6.1 in Agresti saved under `Alligators.txt`.
- The data are from a study of the primary food choices of alligators in four Florida lakes.
- The stomach contents of 219 captured alligators were classified into: Fish, Invertebrate, Reptile, Bird, and Other.

Slide 10

		Primary Food Choice				
Lake	Size	Fish	Inv.	Rept.	Bird	Other
Hancock	small	23	4	2	2	8
	large	4	0	0	1	2
Oklawaha	small	5	11	1	0	3
	large	13	8	6	1	0
Trafford	small	5	11	2	1	5
	large	8	7	6	3	5
George	small	16	19	1	2	3
	large	17	1	0	1	3

Slide 11

$$\begin{aligned}\pi_1 &= \text{prob. of fish,}\\ \pi_2 &= \text{prob. of invertebrates,}\\ \pi_3 &= \text{prob. of reptiles,}\\ \pi_4 &= \text{prob. of birds,}\\ \pi_5 &= \text{prob. of other,}\end{aligned}$$

Let's make "fish" the baseline category. The logit equations are

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_0 + \beta_1 X_1 + \ldots, \quad j = 2, 3, 4, 5.$$

The X's include:

- three dummy variables for lake; we take Hancock as reference level,
- a dummy for size; we take large as reference level.

Slide 12

```
library(VGAM)

# define reference groups
Alligators$Lake = relevel(as.factor(Alligators$Lake), ref="Hancock")
Alligators$Size = relevel(as.factor(Alligators$Size), ref="Large")

# fit multinomial model
gator.fit = vglm(cbind(Invertebrate, Reptile, Bird, Other, Fish) ~
        Lake + Size, family=multinomial, data=Alligators)
```

Slide 13

```
> summary(gator.fit)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept):1    -3.2074     0.6387  -5.021 5.13e-07 ***
(Intercept):2    -2.0718     0.7067  -2.931 0.003373 **
(Intercept):3    -1.3980     0.6085  -2.297 0.021601 *
(Intercept):4    -1.0781     0.4709  -2.289 0.022061 *
LakeGeorge:1      1.6584     0.6129   2.706 0.006813 **
LakeGeorge:2     -1.2428     1.1854  -1.048 0.294466
LakeGeorge:3     -0.6951     0.7813  -0.890 0.373608
LakeGeorge:4     -0.8262     0.5575  -1.482 0.138378
LakeOcklawaha:1   2.5956     0.6597   3.934 8.34e-05 ***
LakeOcklawaha:2   1.2161     0.7860   1.547 0.121824
LakeOcklawaha:3  -1.3483     1.1635  -1.159 0.246529
LakeOcklawaha:4  -0.8205     0.7296  -1.125 0.260713
```

Slide 14

```
LakeTrafford:1    2.7803     0.6712   4.142 3.44e-05 ***
LakeTrafford:2    1.6925     0.7804   2.169 0.030113 *
LakeTrafford:3    0.3926     0.7818   0.502 0.615487
LakeTrafford:4    0.6902     0.5597   1.233 0.217511
Sizesmall:1       1.4582     0.3959   3.683 0.000231 ***
Sizesmall:2      -0.3513     0.5800  -0.606 0.544786
Sizesmall:3      -0.6307     0.6425  -0.982 0.326296
Sizesmall:4       0.3316     0.4482   0.740 0.459506
---
Number of linear predictors:  4
Names of linear predictors:
log(mu[,1]/mu[,5]), log(mu[,2]/mu[,5]), log(mu[,3]/mu[,5]), log(mu[,4]/mu[,5])

Residual deviance: 17.0798 on 12 degrees of freedom
Log-likelihood: -47.5138 on 12 degrees of freedom
Reference group is level  5  of the response
```

Slide 15

Interpretation of parameter estimates

- The intercepts give the estimated log-odds for a particular food choice versus fish in lake=Hancock, size=large.
- For example, the estimated log-odds of birds versus fish for lake Hancock for large alligators (reference group) is -3.2074.
- For the regression coefficients, for instance, the log-odds of invertebrate-versus-fish in Lake George is 1.6584 higher than the log-odds of invertebrate-versus-fish in Lake Hancock, controlling for alligator size.
- In other words, the estimated odds that the primary food choice is invertebrates rather than fish in Lake George are $\exp(1.6584) = 5.25$ times the odds in Lake Hancock, given the same size.

Slide 16

- For example, the prediction equation for the log-odds of selecting invertebrates instead of fish is

$$\begin{aligned}\log\left(\frac{\hat{\pi}_I}{\hat{\pi}_F}\right) &= -3.21 + 1.66\ X_{Geor} + 2.60\ X_{Ockl} \\ &+ 2.78\ X_{Traf} + 1.46 X_{small}.\end{aligned}$$

- For a given lake, for small alligators the estimated odds that the primary food choice is invertebrates rather than fish are $\exp(1.4582) = 4.30$ times the odds for large alligators.
- A Wald 95% CI for the odds that the primary food choice is invertebrates rather than fish for small versus large alligators, controlling for Lake, is

$$\exp(1.4582 \pm 1.96 \times 0.3959) = (1.978, 9.339)$$

Slide 17

- The estimated probability that a large alligator in Lake George has invertebrates as the primary food choice is

$$\hat{\pi}_{IGL} = \frac{\exp(-3.21+1.66)}{1+\exp(-3.21+1.66)+\exp(-2.07-1.24)+\exp(-1.40-0.70)+\exp(-1.08-0.83)} = 0.14$$

Slide 18

The estimated probabilities for each primary food for each of the 8 combinations (size and lake) are:

```
> fitted(gator.fit)
  Invertebrate    Reptile        Bird      Other      Fish
1   0.09309880 0.04745657 0.070401523 0.25373963 0.5353035
2   0.02307168 0.07182461 0.140896287 0.19400964 0.5701978
3   0.60189675 0.07722761 0.008817482 0.05387208 0.2581861
4   0.24864518 0.19483742 0.029416085 0.06866281 0.4584385
5   0.51683852 0.08876722 0.035894709 0.17420051 0.1842990
6   0.19296122 0.20239954 0.108225068 0.20066164 0.2957525
7   0.41285579 0.01156654 0.029671169 0.09380245 0.4521040
8   0.13967784 0.02389871 0.081067366 0.09791362 0.6574425
```

Slide 19

- The saturated model corresponds to the model with interactions between Size and Lake, and has $df = 32$ (4 lakes $\times$ 2 sizes $\times$ 4 logit equations).
- The additive/main effects model has 20 parameters, so its deviance has $df = 32 - 20 = 12$.
- The deviance goodness-of-fit test for the additive model suggests an okay but not great fit to the data

```
> 1-pchisq(17.0798, 12)
[1] 0.1466201
```

Slide 20

Repeating the model-fitting for various sets of predictors, we obtain the following analysis-of-deviance table:

Model	G^2	df	p-value
Saturated**	0.00	0	–
Lake + Size	17.08	12	0.147
Lake	38.17	16	0.001
Size	66.21	24	8.1×10^{-6}
Null	81.36	28	4.2×10^{-7}

** Note: did not converge

The additive model is the best fit we can get for these data.