STA2201H Methods of Applied Statistics II

Monica Alexander

Week 2: Extra notes on categorical outcomes

Categorical data

Categorical data/multinomial responses

- Extension of binomial / binary outcomes.
- Now Y_i make take one of several discrete values, $1, 2, \dots, J$.
- Now the probability is

$$\pi_{ij} = Pr(Y_i = j)$$

with

$$\sum_j \pi_{ij} = 1$$

- As before, for grouped data, n_i is the number of cases in the ith group and y_{ij} is the number of responses that fall in jth category, so the vector of categories $\mathbf{y_i}$ is a of counts that add up to n_i.
- For individual data, $n_i = 1$ and y_{ij} is 0 or 1, so the vector of categories $\mathbf{y_i}$ is a vector of 0s or 1s.

Multinomial distribution

The probability distribution of the counts Y_{ij} given the total n_i is given by the multinomial distribution

$$Pr\{Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}\} = \binom{n_i}{y_{i1}, \ldots, y_{iJ}} \cdot \pi_{i1}^{y_{i1}} \ldots \pi_{ij}^{y_{iJ}}$$

Can this be represented as exponential family?

Conditional distribution

- Let $Y_1, \ldots Y_J$ be Poisson with rate λ_j
- Let $n = \sum Y_j$, which is Poisson with rate $\sum_j \lambda_j$
- Multinomial distribution is joint distribution of Poisson, conditional on sum.

Multinomial regression

- Easy extension to binomial model if we model with respect to a reference category J

$$\eta_{ij} = \log \frac{\pi_{ij}}{\pi_{iJ}} = \mathbf{x_i^T} \beta$$

for $j = 1, \ldots J - 1$.

- Note that if $J = 2$ we have the usual logistic regression
- Coefficients can be interpreted as before, but OR are in relation to reference category

Convert to probabilities

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_k \exp\left(\eta_{ik}\right)} = \mathrm{softmax}(\eta)_i$$

- Choice of reference category would affect βs but not probabilities

Ordered response

What if our categories are ordered? e.g. survey responses are often on an ordinal scale. As before,

$$\pi_{ij} = Pr(Y_i = j)$$

Now consider cumulative probability

$$\gamma_{ij} = Pr(Y_i < j)$$

so

$$\gamma_{ij} = \pi_{i1} + \pi_{i2} + \cdots + \pi_{ij}$$

Model is of the form

$$g(\gamma_{ij}) = \theta_j + \mathbf{x_i^T}\beta$$

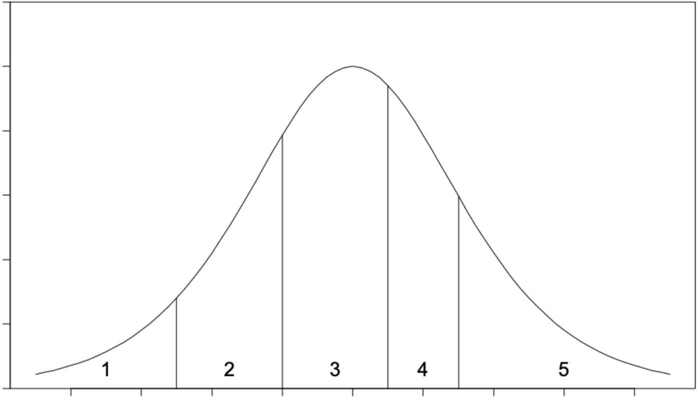
Here θ_j is a constant representing the baseline value of the transformed cumulative probability for category j.

Ordinal regression

Alternatively, can think of a latent variable set-up with cut-points $\theta_1, \ldots, \theta_J$

$$
\begin{aligned}
y_i &= \begin{cases} 1 \text{ if } z_i < \theta_1 \\ 2 \text{ if } z_i \in (\theta_1, \theta_2) \\ \ldots \\ J \text{ if } z_i > \theta_{J-1} \end{cases} \\
z_i &= X_i \beta + \epsilon_i \\
\epsilon_i &\sim f(.)
\end{aligned}
$$

Ordinal regression



Ordinal regression

From the latent formulation

$$\begin{aligned}
\gamma_{ij} &= Pr(Y_i < j) \\
&= Pr(z_i < \theta_j) \\
&= Pr(e_i < \theta_j - X_i\beta) \\
&= F(\theta_j - \mathbf{x_i^T}\beta)
\end{aligned}$$

so

$$g(\gamma_{ij}) = F^{-1}(\theta_j - \mathbf{x_i^T}\beta)$$

as before.

Proportional odds model

Like a logistic regression, but applied to the cumulative probabilities

$$\log \frac{\gamma_{ij}}{1-\gamma_{ij}} = \theta_j + \mathbf{x_i^T}\beta$$

or

$$\frac{\gamma_{ij}}{1-\gamma_{ij}} = \lambda_j \exp(\mathbf{x_i^T}\beta)$$

λ_j is baseline odds of response being in category j.

Pretty strong assumption of proportional odds!

Example

Housing Conditions in Copenhagen

housing	influence	contact	satisfaction	n
tower	low	low	low	21
tower	low	low	medium	21
tower	low	low	high	28
tower	low	high	low	14
tower	low	high	medium	19
tower	low	high	high	37

Example

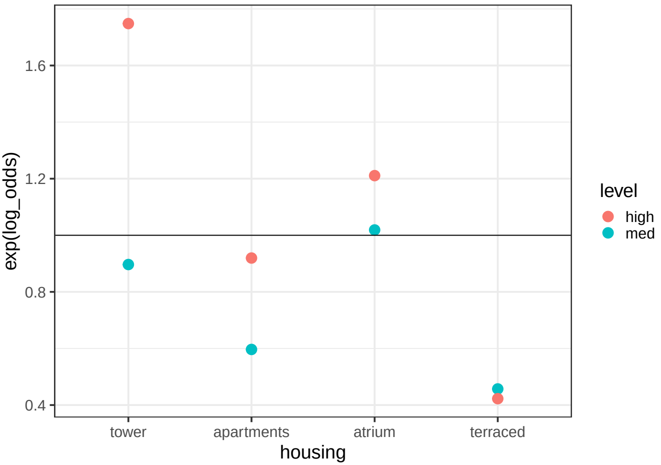
First let's do a multinomial regression, with just housing and contact:

```
## # weights:  18 (10 variable)
## initial  value 1846.767257
## iter  10 value 1793.932058
## final  value 1789.600661
## converged
```

```
## Call:
## nnet::multinom(formula = Y ~ housing + contact, data = copen_wide)
##
## Coefficients:
##            (Intercept) housingapartments housingatrium housingterraced
## sat_medium  -0.1091063         -0.407446     0.1278116      -0.6738718
## sat_high     0.5586042         -0.642400    -0.3672630      -1.4199239
##            contacthigh
## sat_medium   0.3005283
## sat_high     0.3334568
##
## Std. Errors:
##            (Intercept) housingapartments housingatrium housingterraced
## sat_medium   0.1524817         0.1713221     0.2217222       0.2051505
## sat_high     0.1330480         0.1501078     0.2048673       0.1947044
##            contacthigh
## sat_medium   0.1306991
## sat_high     0.1190333
##
## Residual Deviance: 3579.201
## AIC: 3599.201
```

Multinomial regression

Plot the result odds ratios (cf low satisfaction, for low contact)



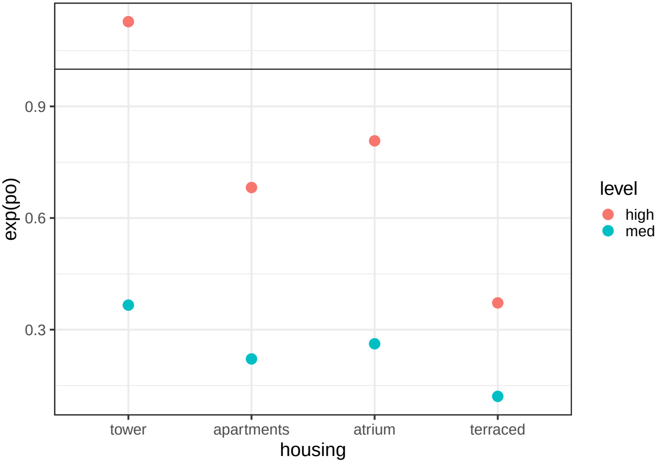
Proportional odds model

Now fit the same idea but with a proportional odds model (ordinal)

```
## Call:
## MASS::polr(formula = satisfaction ~ housing + contact, data = copen,
##     weights = n)
##
## Coefficients:
##                     Value Std. Error t value
## housingapartments -0.5030     0.1169  -4.304
## housingatrium     -0.3341     0.1518  -2.201
## housingterraced   -1.1093     0.1493  -7.428
## contacthigh        0.2540     0.0934   2.720
##
## Intercepts:
##             Value   Std. Error t value
## low|medium  -1.0053  0.1077    -9.3325
## medium|high  0.1202  0.1048     1.1465
##
## Residual Deviance: 3587.389
## AIC: 3599.389
```

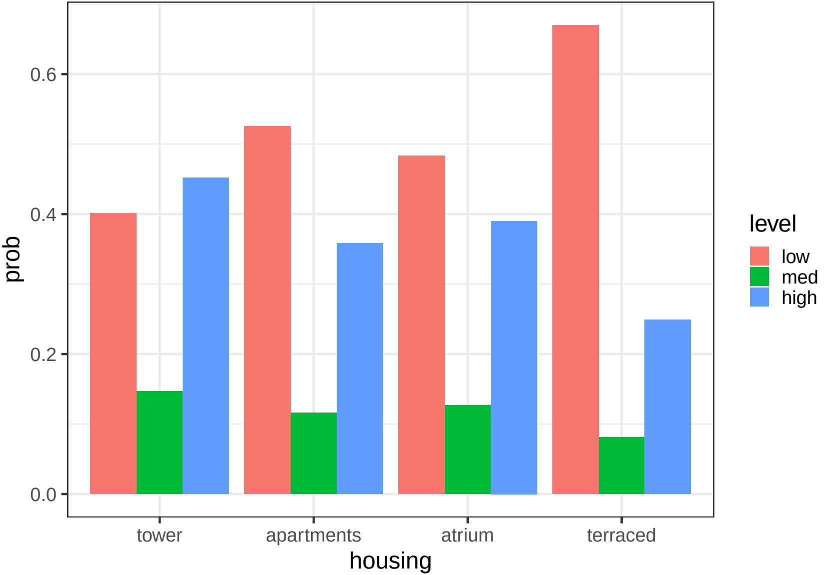
Ordinal regression

Plot of odds ratios



Ordinal regression

Convert log-odds to probabilities:



Summary

- Multinomial models are a natural extension to binomial models
- Looked at logistic forms, but easy to go probit (or other)
- Interpretation is often easiest when we convert to the natural scale