Categorical: Ordinal and multinomial logistic regression

Table of contents

1 Goals **1**
1.1 Goals . . . 1

2 Three or more categories **2**
2.1 Three or more categories . . . 2

3 Multinomial logistic regression **3**
3.1 Multinomial logistic regression . . . 3

4 Ordinal logistic regression **8**
4.1 Ordinal logistic regression . . . 8

5 Summary **14**
5.1 Summary . . . 14

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Extend logistic regression to 3 or more categories
  - Ordered categories (ordinal)
  - Unordered categories (nomimal)
- Probability of "event" (vs "not event")

  - 3+ categories
    - How does that work?

2 Three or more categories

2.1 Three or more categories

2.1.1 Central tendency

- *Nominal*: **Unordered** categories
  - Central tendency: **Mode**
    - Most common response
- *Ordinal*: **Ordered** categories
  - Central tendency: **Median**
    - 50th percentile: Half of responses higher, half lower

2.1.2 Multiple categories

- Won't deal with median or mode
  - Instead: *Probability* of each category
  - *Compare* different probabilities
- **What do we compare?**
  - Everything to everything?
  - Everything to one category?
  - Something else?

2.1.3 Multinomial and ordinal logistic regression

- **Multinomial logistic regression**
  - *No order* to categories: Vanilla, chocolate, strawberry
  - Compare all categories to a **reference category**
- **Ordinal logistic regression**
  - Categories are *ordered*: Disagree → Neutral → Agree
  - Compare each category to **all higher** (or **all lower**, depending on software) categories

3 Multinomial logistic regression

3.1 Multinomial logistic regression

3.1.1 Multinomial logistic regression

- **Outcome**: Nominal
  - *Department*: Psychology, Epidemiology, Statistics, Business
  - *Religion*: Christian, Jewish, Muslim
  - *Ice cream flavor*: vanilla, chocolate, strawberry
- **Distribution**: Multinomial (generalized binomial to more than 2)
- **Link function**: Logit

3.1.2 Multiple equations

- Multiple equations for this model
  - With a categories, you have $(a-1)$ equations
  - Actually same as logistic: 2 categories $\rightarrow 2-1=1$ equation
- One category is "reference" category
  - All other categories are compared to the reference category
  - A bit like dummy codes, but outcome instead of predictor

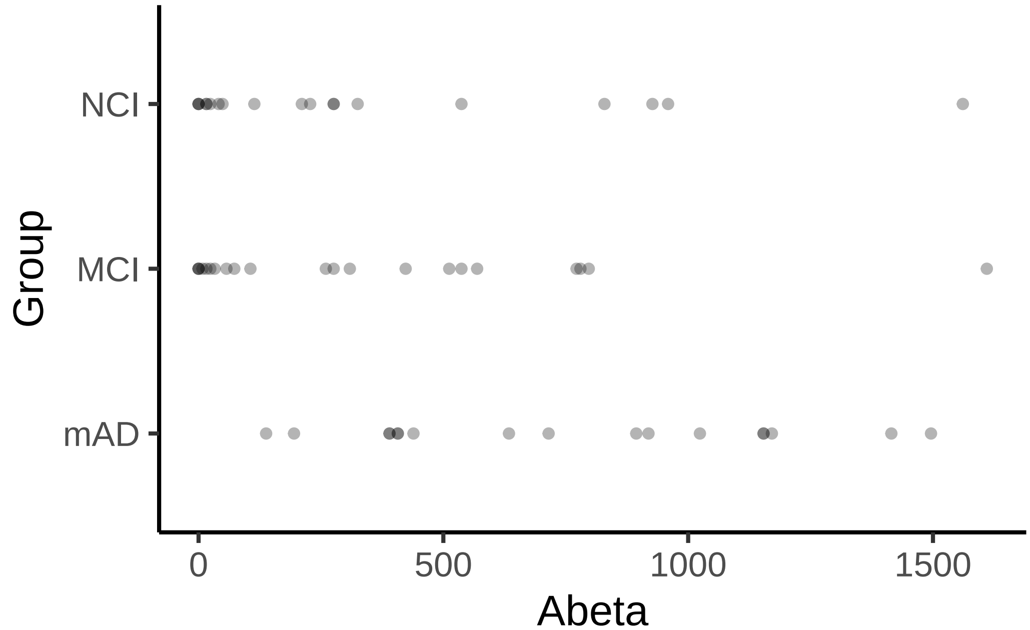
3.1.3 Example: Multinomial logistic regression

- `Amyloid` dataset in **Stat2Data** package
  - `Abeta`: Amyloid-β in posterior cingulate cortex (pmol/g tissue)
  - `Group`:
    - `mAD` = Alzheimer's disease
    - `MCI` = mild cognitive impairment
    - `NCI` = no cognitive impairment

3.1.4 Frequencies: Amyloid dataset

Var1	Freq	Prob
mAD	17	0.298
MCI	21	0.368
NCI	19	0.333

3.1.5 Figure: Amyloid dataset



3.1.6 Multiple equations

- *3 groups*: `mAD`, `MCI`, `NCI`
  - $a = 3 \rightarrow a - 1 = 2$ *equations*
    - `mAD` is **reference**
    - *Equation 1*: `MCI` vs `mAD`
    - *Equation 2*: `NCI` vs `mAD`
- Mutually exclusive, so $p_{m_A D} + p_{MCI} + p_{NCI} = 1$
  - Everyone is in exactly 1 `Group`
  - Never anything else, never multiple options

3.1.7 Multinomial logistic regression equations

$$ln\left(\frac{\hat{p}_{MCI}}{\hat{p}_{mAD}}\right) = b_{0,2.1} + b_{1,2.1}Abeta$$

$$ln\left(\frac{\hat{p}_{NCI}}{\hat{p}_{mAD}}\right) = b_{0,3.1} + b_{1,3.1}Abeta$$

- Notation: predictor, numerator category, denominator category
  - **All coefficients are different across equations**

3.1.8 Example: Output

```
# weights:  9 (4 variable)
initial  value 62.620900
final  value 57.318323
converged

Call:
multinom(formula = Group ~ Abeta, data = Amyloid)

Coefficients:
    (Intercept)        Abeta
MCI    1.319761 -0.002092564
NCI    1.231750 -0.002128210

Std. Errors:
    (Intercept)        Abeta
MCI   0.5583894 0.0008282646
NCI   0.5666824 0.0008558914

Residual Deviance: 114.6366
AIC: 122.6366
```

3.1.9 Example: Coefficients and p-values

- Coefficients for the model

	(Intercept)	Abeta
MCI	1.320	-0.002
NCI	1.232	-0.002

- p-values for the model

	(Intercept)	Abeta
MCI	0.018	0.012
NCI	0.030	0.013

3.1.10 Example: Confidence intervals

	2.5 %.MCI	97.5 %.MCI	2.5 %.NCI	97.5 %.NCI
(Intercept)	0.225	2.414	0.121	2.342
Abeta	-0.004	0.000	-0.004	0.000

3.1.11 Example: Multinomial logistic regression equations

$$ln\left(\frac{\hat{p}_{MCI}}{\hat{p}_{mAD}}\right) = b_{0,2.1} + b_{1,2.1}Abeta = 1.32 + (-0.00209)Abeta$$

$$ln\left(\frac{\hat{p}_{NCI}}{\hat{p}_{mAD}}\right) = b_{0,3.1} + b_{1,3.1}Abeta = 1.232 + (-0.00213)Abeta$$

- Slopes **are not** the same across the two equations

3.1.12 Example: Multinomial logistic regression equations

- One regression coefficient **per predictor**, *per equation*
  - `Abeta` has a certain effect on the probability of having **mild impairment** vs **Alzheimer's** ($b_{1,2.1}$)
  - `Abeta` has a **different** effect on the probability of having **no impairment** vs **Alzheimer's** ($b_{1,3.1}$)
  - Here, the values are **very** close

3.1.13 Example: Interpretation

- Interpret as in (binary) logistic regression, **except**
  - Binary logistic regression: "success" vs "not success"
  - Here: **Numerator category** vs **denominator category**
    - Subset of total number of categories

3.1.14 Example: Interpretation

- Odds interpretation of intercept
  - Odds of `MCI` vs `mAD` $= e^{b_{0,2.1}} = e^{1.32} = 3.743$
    - `Abeta` $= 0$: Odds of `MCI` is 3.743 times higher than odds of `mAD`
      - With no amlyoid-β, you're much more likely to have *mild impairment* than *Alzheimer's*
  - Odds of `NCI` vs `mAD` $= e^{b_{0,3.1}} = e^{1.232} = 3.427$
    - `Abeta` $= 0$: Odds of `NCI` is 3.427 times higher than odds of `mAD`
      - With no amlyoid-β, you're much more likely to have *no impairment* than *Alzheimer's*

3.1.15 Example: Interpretation

- Odds interpretation of effect of `Abeta`
  - Odds ratio for `MCI` vs `mAD` $= e^{b_{1,2.1}} = e^{-0.0020926} = 0.99791$
    - < 1: **More `Abeta`** means **lower odds of `MCI`** (*relative to `mAD`*)
      - More amyloid-β means *more likely to have Alzheimer's*
  - Odds ratio for `NCI` vs `mAD` $= e^{b_{1,3.1}} = e^{-0.0021282} = 0.99787$
    - < 1: **More `Abeta`** means **lower odds of `NCI`** (*relative to `mAD`*)
      - More amyloid-β means *more likely to have Alzheimer's*

3.1.16 Important note 1

Warning

- With 3 categories, there are **3 possible comparisons**
  - The third comparison is *redundant* (similar to dummy codes)
    - But we can calculate it
      - $b_{1,2.3} = b_{1,2.1} - b_{1,3.1}$
      - Or re-order the outcome and re-run

3.1.17 Important note 2

Warning

- Most statistical presentations: **Last** category as reference
  - SPSS and SAS: **Last** category as the reference (default)
  - R (and here): **First** category as the reference
    - How is it different?
      - You'll get the "missing" third comparison instead
      - Some signs will flip because you're making the opposite order comparison: $\frac{\hat{p}_{MCI}}{\hat{p}_{mAD}}$ vs $\frac{\hat{p}_{mAD}}{\hat{p}_{MCI}}$

3.1.18 Some difficulties

- There are **many** regression coefficients to interpret
  - For 3 outcome categories and 1 predictor
    - 4 coefficients to interpret
  - More coefficients with more predictors
    - $(a-1)$ more coefficients for each added predictor

4 Ordinal logistic regression

4.1 Ordinal logistic regression

4.1.1 Ordered categorical outcomes

- Outcome categories have a **natural ordering or progression**
  - Make some *simplifications* to multinomial logistic regression model
- Ordinal logistic regression model is
  - **Much easier to interpret**
  - **Better power**
  - **A few additional assumptions**

4.1.2 Ordinal logistic regression

- **Outcome**: Ordinal
  - Dose of treatment: low, medium, high
  - Rankings: 1st, 2nd, 3rd, 4th
  - Education: high school, some college, college grad, graduate
  - Likert scales: agree, neutral, disagree
- **Distribution**: Binomial
- **Link function**: Cumulative logit
  - This model is also called the "cumulative logit model"

4.1.3 Multiple equations

- Multiple equations for this model
  - With a categories, you have $(a-1)$ equations
- Take advantage of the **ordering** of categories
  - Category 1 then category 2 then category 3
    * Category 1 vs all higher
    * Categories 1 and 2 vs all higher

4.1.4 Multiple equations

- *3 groups*: `mAD`, `MCI`, `NCI`
  - $a = 3 \rightarrow a - 1 = 2$ equations
    * Ordered: `mAD` then `MCI` then `NCI`
    * *Equation 1*: `mAD` vs all higher
    * *Equation 2*: `mAD` and `MCI` vs all higher
- Mutually exclusive, so $p_{m_A D} + p_{MCI} + p_{NCI} = 1$
  - Everyone is in exactly 1 `Group`
  - Never anything else, never multiple options

4.1.5 Ordinal logistic regression equations

$$ln\left(\frac{\hat{p}_{mAD}}{\hat{p}_{MCI}+\hat{p}_{NCI}}\right) = b_{0,1} + -b_1 Abeta$$

$$ln\left(\frac{\hat{p}_{mAD}+\hat{p}_{MCI}}{\hat{p}_{NCI}}\right) = b_{0,12} + -b_1 Abeta$$

- **All slopes are the same across equations**
  - Intercepts are still different

4.1.6 Example: Output

```
Call:
polr(formula = Group ~ Abeta, data = Amyloid, Hess = TRUE)

Coefficients:
          Value Std. Error t value
Abeta -0.001671  0.0006333  -2.639

Intercepts:
        Value   Std. Error t value
mAD|MCI -1.6689  0.4323    -3.8602
MCI|NCI  0.0729  0.3618     0.2014

Residual Deviance: 116.6483
AIC: 122.6483
```

4.1.7 Example: Coefficients and p-values

	Value	Std. Error	t value	p value
Abeta	-0.002	0.001	-2.639	0.008
mAD|MCI	-1.669	0.432	-3.860	0.000
MCI|NCI	0.073	0.362	0.201	0.840

4.1.8 Example: Confidence intervals

- Only get CIs for the **slope**, not intercepts

```
       2.5 %        97.5 %
-0.002961670 -0.000509697
```

4.1.9 Important note 1

> ⚠ Warning
>
> - Remember in logistic regression when I mentioned that the model is sometimes presented as
>   - $\hat{p} = \frac{1}{1+e^{-(b_0+b_1X)}}$
>   - With a negative sign?
> - Ordinal logistic regression in R does a *similar* thing
>   - Use the **negative** of the **slope(s)** for interpretation
>   - All metrics
> - SPSS and SAS have their own weird approaches to this
>   - Results do not match across R, SPSS, SAS

4.1.10 Example: Ordinal logistic regression equations

$$ln\left(\frac{\hat{p}_{mAD}}{\hat{p}_{MCI}+\hat{p}_{NCI}}\right) = b_{0,1} + -b_1 Abeta = 1.669 + (0.002)Abeta$$

$$ln\left(\frac{\hat{p}_{mAD}+\hat{p}_{MCI}}{\hat{p}_{NCI}}\right) = b_{0,12} + -b_1 Abeta = -0.073 + (0.002)Abeta$$

- Slopes **are** the same across the two equations

4.1.11 Example: Ordinal logistic regression equations

- One regression coefficient **per predictor**
  - `Abeta` has a certain effect on the probability of having Alzheimer's vs (mild or no cognitive impairment) (b_1)
  - `Abeta` has the **same effect** on the probability of having (Alzheimer's or mild cognitive impairment) vs no cognitive impairment (b_1)
- This assumption is called the **proportional odds assumption**
  - A predictor has the same effect on **changing categories** regardless of *which* categories you are switching between

4.1.12 Example: Interpretation

- Interpret as in (binary) logistic regression, **except**
  - Binary logistic regression: “success” vs “not success”
  - Here: **Numerator category or categories** vs **denominator category or categories**
    - All categories

4.1.13 Example: Interpretation

- Odds interpretation of intercept
  - Odds of `mAD` vs (`MCI` and `NCI`) $= e^{b_{0,1}} = e^{-1.669} = 0.188$
    - `Abeta` $= 0$: Odds of `mAD` is 0.188 times odds of `MCI` and `NCI`
      - With no amlyoid-β, you’re less likely to have *Alzheimer’s* than (*mild impairment* or *no impairment*)
  - Odds of (`mAD` and `MCI`) vs `NCI` $= e^{b_{0,12}} = e^{0.073} = 1.076$
    - `Abeta` $= 0$: Odds of `mAD` and `MCI` is 1.076 times odds of `NCI`
      - With no amlyoid-β, you’re more likely to have (*Alzheimer’s* or *mild impairment*) than *no impairment*

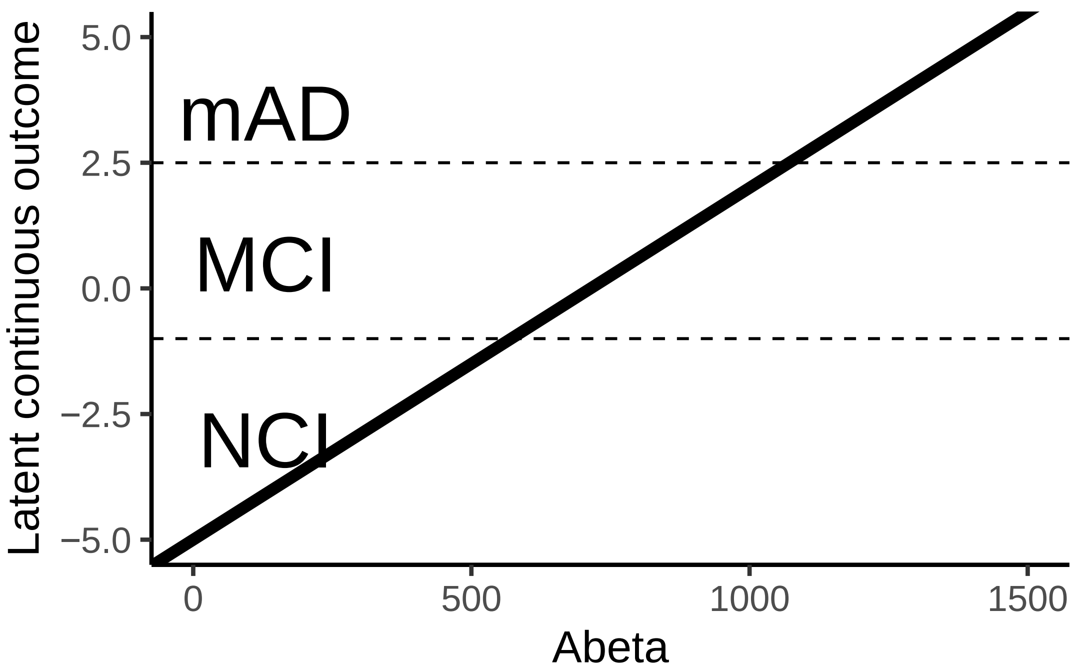
4.1.14 Example: Interpretation

- Odds interpretation of effect of `Abeta`
  - Odds ratio for `Abeta`: $e^{-b_1} = e^{0.002} = 1.002$
    - > 1: **More `Abeta`** means
      - **Higher odds of `mAD`** *relative to (`MCI` and `NCI`)*
      - More amyloid-β means *more likely to have Alzheimer’s*
      - **Higher odds of (`mAD` and `MCI`)** *relative to `NCI`*
      - More amyloid-β means *more likely to have Alzheimer’s or mild impairment*

4.1.15 Important note 2

> ⚠ Warning
>
> - Note that **R** orders the outcome categories in **alphabetical** order by default
>   - Just *happens* to corresponds to highest to lowest severity in this example
>   - If that's not true in your dataset
>   - Manually re-order levels (e.g., **forcats**)
>   - Recode the outcome to numbers with the correct order

4.1.16 Figure: Proportional odds (conceptual, not to scale)



4.1.17 Proportional odds

- The slope (e.g., b_1) is the same regardless of going from `mAD` to `MCI` or from `MCI` to `NCI`
  - Regardless of which "threshold" you are crossing
- Proportional odds *simplifies* things compared to the multinomial logistic regression model
  - Fewer coefficients

- Predictors have the same effect on changing categories regardless of which categories

4.1.18 Testing proportional odds

- Manually split outcome into
  - `mAD` vs all higher
  - `mAD` and `MCI` vs all higher
    - Run logistic regression on each
    - If proportional odds holds, **slopes** in both models are very close

term	estimate	estimate
Abeta	0.002	0.001

4.1.19 Testing proportional odds

- An easier but "not completely statistically correct" approach (Hosmer & Lemeshow, page 304)
  - Likelihood ratio test comparing the *multinomial* and *ordinal* logistic regression models
  - $\chi^2(1) = 2.012, p = 0.156$
    - Test is NS, so use the simpler model (*ordinal*)

5 Summary

5.1 Summary

5.1.1 Summary of this week

- Extend binary logistic regression to 3+ categories
  - Unordered = Multinomial logistic regression
    - A LOT of coefficients to estimate
    - Reference category
  - Ordered = Ordinal logistic regression
    - Simpler model with fewer coefficients
    - Proportional odds assumption

5.1.2 Next week

- Models for count outcomes
  - Poisson regression
  - Overdispersed Poisson regression
  - Negative binomial regression
  - Excess zeroes versions of these models