Multilevel analysis

Dichotomous and ordinal data

Emmeke Aarts (E.Aarts@uu.nl)
Methodology and Statistics
Faculty of Social and Behavioral Sciences

1

Course outline

Lecture 1
- when/why multilevel analysis
- the multilevel regression model

Lecture 2
- Longitudinal model
- Contextual effects

Lecture 3
- Analyzing dichotomous and ordinal data

2

Today

1) Generalized linear models
2) Multilevel analysis for dichotomous data
3) An example
4) Multilevel analysis for ordinal data
5) An example

Generalized linear models

Modelling non-continuous outcomes

<count>
poisson-distributed variable ~ EX) # of car accident within certain time frame

Generalized linear model

- Standard linear regression: continuous outcome variable with normal distribution for errors
- Some outcome variables *must* violate this assumption
  - Dichotomous outcome variable
  - Ordinal outcome variable
  - Nominal outcome variable
  - Count outcome variable

For all of these, we use the generalized linear framework.
How this works: transform the outcome variable such that it does have a linear function in the predictors.
we use tricks so we can use linear model to predict the outcomes

<ordinal>
more than 2 categories + ranking in categories
EX) bad - Soso - good - very good

<nominal>
categories w/o order.
EX) What is your political preferences, blood type ..

EX) logistic regression, ordinal logistic regression
probit regression, ... poisson regression ...

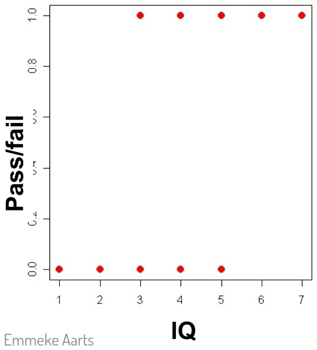
Generalized linear model

- Solution: generalized linear model (GLM)
- Transform the outcome such that it becomes a linear function of the predictors
- Examples
  - Logistic regression,
  - Ordinal logistic regression
  - Probit regression
  - Poisson regression

problem using linear model for binary outcomes
values greater than 1 & smaller than 0...

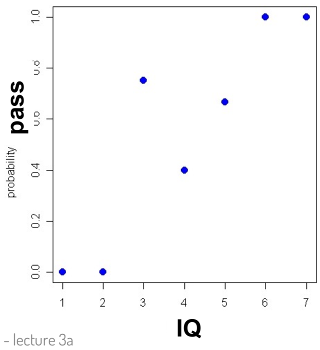
Logistic regression

Problem:
Outcome can only be 0 or 1; Do we want to draw a straight line through that?...



Solution:
Rephrase problem: predict 'probability to 0 or 1' rather 0/1 itself

Still problem: we'll still predict prob. smaller than 0, & larger than 1.



Emmeke Aarts Multilevel analysis - lecture 3a 7

7

Logistic regression

we're still predicting the prob. of success but instead, we're using transformation here: predicting $\frac{e^x}{1+e^x}$

linear function

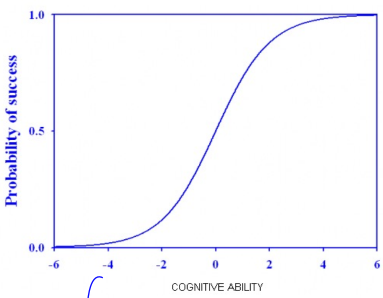
That way, our outcomes are bounded between 0 & 1

Probability of event Y=1, P(Y=1), is often expressed in terms of a logistic function:

$$P(Y=1) = \frac{e^x}{1+e^x}$$

whatever we fill in X (our predictors) it cannot become <0 or >1

Advantages of this function
- Outcome is between 0 and 1
- Monotone increase (or decrease)



outcome is bounded between 0 & 1

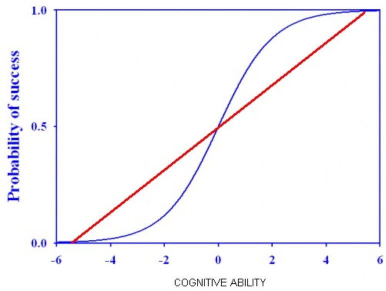
Emmeke Aarts Multilevel analysis - lecture 3a 8

8

Logistic regression

Last problem: The function we use for the expression of probability is *non*-linear and we want to use linear regression, which assumes a linear relation between the predictor variable and the outcome variable…

Solution: "logarithmic transformation"



Logistic regression

Remember: $\ln(e^x) = x$

$$\text{odds} = \frac{P(Y=1)}{P(Y=0)} = \frac{p_i}{1-p_i} = e^{(\beta_0 + \beta_1 x_i + \cdots)} = \frac{\text{prob. success}}{\text{prob. failure}}$$

$$\ln(\text{odds}) = \beta_0 + \beta_1 x_i + \cdots$$

So if we take the log of the odds, and use this this measure as the outcome variable, we can use the linear regression model

Important to remember: when using logistic regression, we do not predict 0/1 outcomes, but the "log of the odds," i.e., the log of the ratio of the probability of event 1 vs probability of event 0

→ If we're using logistic reg., what we're predicting is "log of odds"
so we donot predict 0, 1 outcomes, nor probabilities.

⇓

your reg. coefficients (βs) are related to the log of odds → log of the ratio of prob. that your event = 0 vs event = 1

Probability versus odds: Lucky box #1

Lucky box #1
6 blue
3 red
2 yellow
1 green

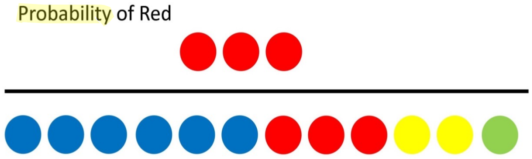
red = win

How to express chances of winning? ~> 2 ways { prob. of winning / odds of winning

11

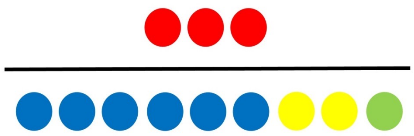
Probability versus odds: Lucky box #1

Probability of Red



$$P(red) = \frac{number\ red}{total\ number}$$

Odds for Red



$$odds(red) = \frac{number\ red}{number\ not\ red}$$

12

Probability versus odds: Lucky box #2

Lucky box #2
5 blue
4 red
2 yellow
1 green

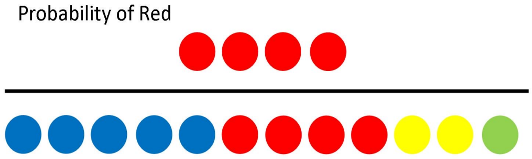
red = win

How to express chances of winning?

13

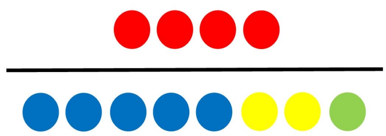
Probability versus odds: Lucky box #2

Probability of Red



$$P(red) = \frac{number\ red}{total\ number}$$

Odds for Red



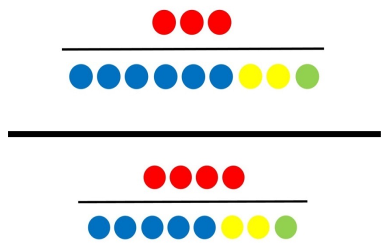
$$odds(red) = \frac{number\ red}{number\ not\ red}$$

14

Odds ratio = ratio of ratios!

What does it mean? lower odds of winning in box 1 compared to box 2.

$$odds\ ratio = \frac{odds(red)\ lucky\ box\ 1}{odds(red)\ lucky\ box\ 2} =$$



Odds ratio is a ratio of ratios.
Odds ratio can never be negative.

DIY: what is the odds ratio of winning in box 1 over winning in box 2?

$$\frac{\frac{3}{9}}{\frac{4}{8}} = \frac{2}{3} < 1,\ \text{smaller than } 1$$

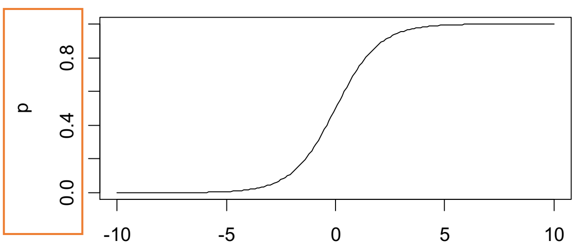
▶ Another way to interpret is: winning in box 1 is 0.33 ($\frac{1}{3}$) less likely, compared to box 2.

"the chances of winning in box 1 is smaller than winning in box 2."

Logit transformation

- For proportions the logit transformation is
$$f(p)=logit(p)=\ln(p/(1-p))$$
- Nonlinear transformation that maps the probability scale $p \in [0,1]$ on the linear predictor $\eta \in [-\infty,+\infty]$. unbounded

Latent, we do not observe p !!



η ~ latent construct: vehicle we're using that enables us to use this linear model

Logistic regression

- Various ways to formulate logistic regression model
  - in terms of the probability: $p_i = \frac{e^{(\beta_0+\beta_1 x_i+\cdots)}}{1+e^{(\beta_0+\beta_1 x_i+\cdots)}}$ → compute predicted prob. of success given observed values in predictors
  - in terms of the odds: $\text{odds} = \frac{p_i}{1-p_i} = e^{(\beta_0+\beta_1 x_i+\cdots)}$ = $\frac{\text{prob. of success}}{\text{prob. of failure}}$
  - in terms of the logit: $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i + \cdots$ = log(odds) → If you just look at the raw reg. coef. then, what you're looking at is log of the odds.
- Note there is no error variance since the residual variance is known: $\text{var}(y) = p_i\,(1-p_i)$ fixed in logistic regression so it's not given in the output.
- Furthermore, expected outcome: $E(y) = p_i$ is the prob. of success

Logistic regression

we don't interpret the beta coefficients directly, but we have various options.

- Interpretation regression coefficients $\beta_1, \beta_2, \ldots$
  - qualitatively: positive or negative effect on logit ⇒ just look at the signs: Do they increase / decrease the expected chance of success
  - quantitatively:
    - effect on logit is β : but this isn't very informative.
    - effect on odds is $\exp(\beta)$ ⇒ expected proportion of change in odds (by taking exponent, we get what is the effect is on odds of the particular variable)
  - latent variable interpretation
- Interpretation intercept β_0: as threshold θ for latent variable interpretation of logistic regression

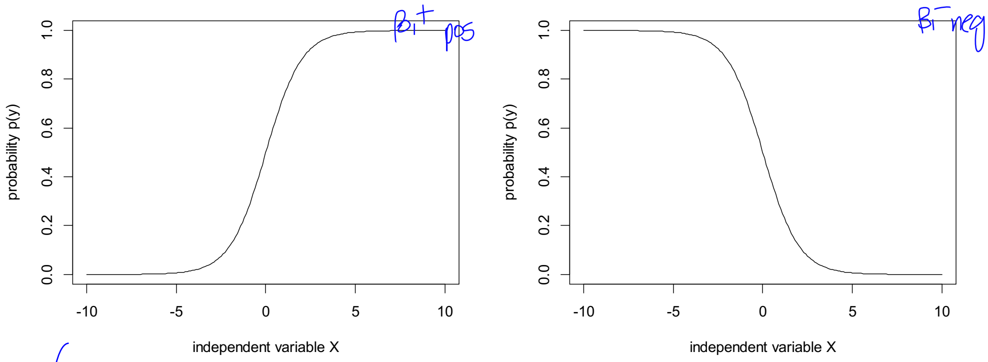
So if you've centered the variables, for ex, then when all our predictors are average, what is the predicted prob. of success? ∴ β_0

Interpretation of weights β_0 and β_1

predicted chance of success when all the covariates = 0

β_0 determines p(y) when X=0

β_1 is positive left and negative right

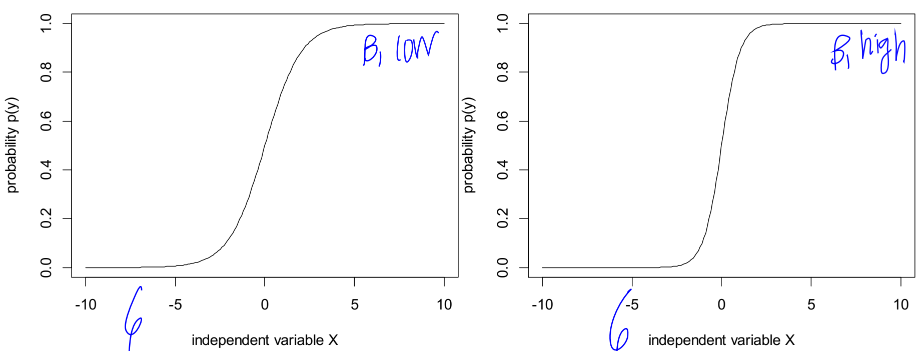


EX) X = IQ, y = passing a test

Then, as IQ increases, the prob. of passing the test also increases.

Interpretation of weight β_1

size β_1 determines the steepness of the logistic curve



meaning that it takes quite a lot of increases in X to get the prob. of 0 to prob. of 1

if β_1 is really high, it's almost like a switch: w/ marginal increase in X, we rise to the prob. of 1 very quickly.

Coffee break

21

Eta is the thing we're predicting by the log of odds

Latent variable presentation

Distribution of our sample over η

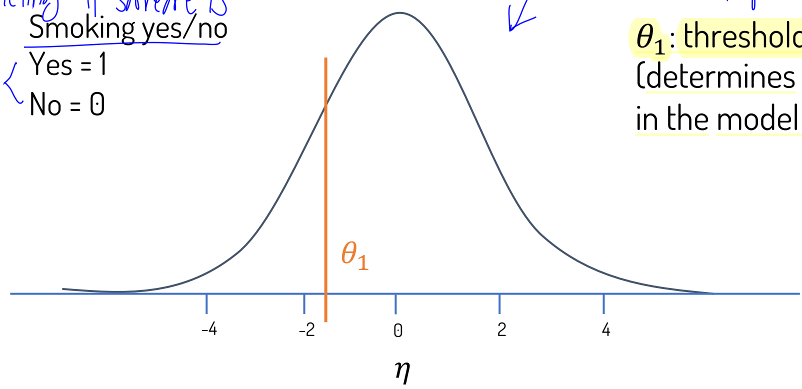
This is dist. of predicted η values in our samples by using our model

we're predicting if someone is

Smoking yes/no

Yes = 1

No = 0



θ_1: threshold, equeal to β_0 (determines p(y) when all predictors in the model are X=0)

Remember: η = ln(p / [1 – p])

22

θ₁ is negative → what does it tell you about our sample?

If $p<0.5$, then $\ln\left(\frac{p}{1-p}\right)$ will become neg.

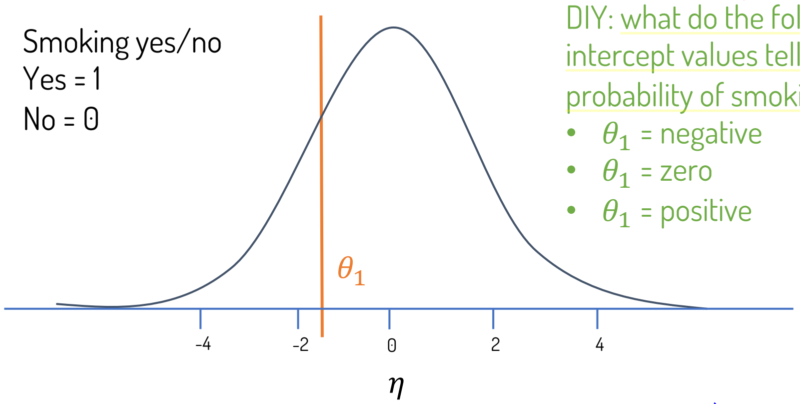
3/3/22

$\theta_1=0$ → $p=0.5$: means in our sample, if all our predictors are zero, then the predicted ratio of ppl who smoke in our data is 0.5

— If we now center our continuous variables, then just running a model & having a glance at the intercept, we immediately know if for the average ppl in our sample, whether we expect the large proportion of them to pass or not pass.

Latent variable presentation

Smoking yes/no
Yes = 1
No = 0



DIY: what do the following threshold / intercept values tell you about the baseline probability of smoking:

- θ_1 = negative
- θ_1 = zero
- θ_1 = positive

Ex) we have a continuous var & dichotomous var: sex (boy = 0, girl = 1) then our threshold will tell us: what proportion do we predict the boys to pass/smoke? (given that continuous predictors're centered)

↓ we just need to see the sign of the intercept! Then we know

neg: proportion is smaller than 0.5
0: proportion = 0.5
pos: proportion is larger than 0.5
↓ majority of samples pass!

Remember: $\eta = \ln(p / [1 - p])$

※ NOTE: what you code it as "success": what represents 1 in your dataset,

Latent variable presentation

Latent variable $Y_i = \beta_0 + \beta_1 x_i + \cdots + e_i$

Latent variable is unobserved: its scale (η) is arbitrary and to identify the model it needs to be standardized. !!

fixed variance

Standard logistic distribution: $mean = 0$ and variance $= \pi^2/3$ (= 3.29)

Standard normal distribution: $mean = 0$ and variance $= 1$

we don't have the lowest residual variance as an outcome, becuz it's fixed as $\frac{\pi^2}{3}$ (= 3.29) It's always 3.29.

Model building:

- β's change because of adding predictors
- β's also change because of rescaling

Take home msg

Be aware! they're re-scaled → they cannot be compared over models as you used to do w/ normal linear regression ML-analysis

This has a consequence tho.

Becuz ofc, it's a bit strange that, no matter which model you fit, no matter which predictor you include the residual var. is always 3.29! That's odd! becuz we're adding predictors, we're explaining more things.. Well, the βs change becuz of adding predictors, but they also change between diff. models becuz of the rescaling, which means if you're using multi-level logistic regression, you cannot compare βs w/ subsequent models!

Becuz they are on a diff. scale. Becuz of this rescaling, becuz it always have to fit this residual var. as 3.29.
So the whole thing is stretched ...
To fit this standard logistic distribution, which means we cannot compare

βs over subsequent models.

Multilevel binary logistic model

Binary outcomes

Regression equation

- At the lowest level not $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$
- but $\text{logit}(p_{ij}) = \beta_{0j} + \beta_{1j}x_{ij}$ (no error term)
- Then, as usual:
  - $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$
  - $\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$
- combining (substitution and rearranging terms) gives
  - $\text{logit}(p_{ij}) = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij}$

now we're predicting the logit (p) = log (odds) for prob of success w/o error term, becuz it's 3.29!

at level 2, we're predicting cluster-dependent intercept by overall intercept + second level residual var. for intercept + level 2 predictor which partly explain why these intercepts differ over the clusters & the same idea for cluster-specific slopes: overall slope + residual deviation to the overall slope + possibly cross-level interaction that might explain why the effect of the predictor differs over the clusters

mixed equations (no residual error term at the lowest level: e_{ij}) & we're here not modeling Y_{ij} directly! we're modeling the log (odds)

Explained variance

we can look at the explained var. of eta, η, does have the residual!

- Assume latent variable Y_{ij} (η), *this variable does have residual*
- Multilevel model: $Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + e_{ij}$

- Variance of first level error term is
  - $var(e_{ij}) = \sigma_R^2 = \pi^2/3$ for logistic regression
  - $var(e_{ij}) = \sigma_R^2 = 1$ for probit regression

is fixed as 3.29, but there is still a way to calculate R^2: we still have second-level error term (intercept variance)

- Variance of second level error term is $var(u_{0j}) = \sigma_{u0}^2$

And: variance of linear predictors

this is new ★ "Variance of linear predictor" from the fixed part is $\sigma_F^2 = var(\gamma_{00} + \gamma_{10}x_{ij})$

tells us about the variance in the model that we can explain in the fixed part of the model, due to the fixed part of the regression.

we're predicting η using the observed covariate values that we have in our data.

You could see this in such a way that diff. in this prediction, that's the diff. between what we're explaining w/ fixed effects and what we have in our data.

So we can calculate the variance of linear predictor by calculating the variance over this predicted η value using reg. coef. that we get from the model output.

Explained variance

Proportion explained variance: $R_{MZ}^2 = \frac{\sigma_F^2}{\sigma_F^2 + \sigma_{u0}^2 + \sigma_R^2}$

$= \frac{\text{var. of linear model}}{\text{total variance}}$

- Residual variance
  - $\sigma_R^2 = \frac{\pi^2}{3} \approx 3.29$ for logistic regression
  - $\sigma_R^2 = 1$ for probit regression

- Second-level variance σ_{u0}^2 from multilevel software output (e.g., R)

- Variance of linear predictor of fixed part σ_F^2 is calculated in R using estimates of regression coefficients

sth. that you need to calculate yourself!

Variance of the linear predictor

var. of latent factor we calculated
that shows us how much do we explain of all the differences on our latent construct by using the fixed part of model!

ID	sex	age	...	$\gamma_{00} + \gamma_{10}\boldsymbol{sex}_{ij} + \gamma_{10}\boldsymbol{age}_{ij} +$ [redacted] $_{ij}$
1	0	23		-2.8
2	0	20		-2.71
3	1	19		-2.27
4	1	25		-2.45
5	1	24		-2.42
6	0	28		-2.95
7	1	22		-2.36

$\sigma_F^2 =$ <- var

Then this tells us how much differences on our latent scale we explain by our fixed regression coefficients.

then we ask for the variance of these values $= \sigma_F^2$

$Y_{ij} = \gamma_{00} + \gamma_{10}\boldsymbol{sex}_{ij} + \gamma_{20}\boldsymbol{age}_{ij}$ fixed part of the model

$Y_{ij} = -2.11 + 0.41\ \boldsymbol{sex}_{ij} - 0.03\ \boldsymbol{age}_{ij}$ → calculate predicted η value using this equation & the observed value for this person & we do this for everyone.

Using this linear function, we calculate the predicted value of η (eta) for everyone in the dataset.

Intraclass correlation coefficient

Again look at η, becuz it has residual!

- Assume latent variable Y_{ij}, *this variable does have residual*
- Empty multilevel model: $Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$ ~ use intercept-only model then we still get our intercept variance

- $\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_R^2}$ as usual

instead of σ_e^2
now here we use 3.29 ~ That's the only difference.

Estimation

- Likelihood function includes a difficult integral over the random effects , which means that this estimating parameters is tricky.
- Estimating parameters in multilevel generalized linear models is a problem
- Solutions
  - Linearization of the likelihood function by means of Taylor series expansion: MQL/PQL methods (likelihood estimates and so deviance and AIC unreliable)
    - MLwiN, HLM
  - Numerical integration of the likelihood function (likelihood estimates ok)
    - R, Supermix, Mplus, HLM (dichotomous outcomes)
  - Bayesian methods
    - MLwiN, Mplus

Every software package uses diff. numerical optimization

↓

outcome can be really diff. depending on what software you use.

Coffee break

An example

Of logistic multilevel regression on dichotomous outcomes

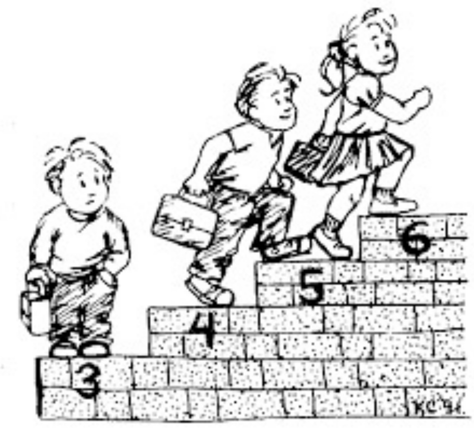
33

Thailand Education Data

- 7516 pupils in primary schools
  - Example data from HLM
- Outcome variable dichotomous
  - Repeat a class (1) or not (0)
- level 1 • Pupil level predictors
  - sex (1=male, 0=female)
  - pre-primary education (PPED: 1=yes, 0=no)

nested ∧

- level2 • School level predictor
  - Mean SES



we're predicting: Do they have to repeat a class or not?

34

① Empty model: neg. intercept (-2.23) → majority of the children does not have to repeat.

② Introducing level 1 predictors: intercept doesn't change much.

- Sex: If you're a boy (1) instead of a girl (0), we have a pos. value → meaning that if you're a boy, you're more likely to repeat a class cuz the reg. coef is pos.
- PPED: reg. coef. is negative → meaning that if you went to PPED, you're less likely to repeat a class.

3/3/22

Thailand Education Data

Model	Empty model ① intercept only model		② Predictors level 1, fixed effects		
Predictor	par est	SE	par est	SE	Odds ratio
Fixed part					
threshold	-2.23	0.08	-2.24	0.10	
Sex (female=ref)			0.54	0.07	1.71
PPED (no=ref)			-0.64	0.10	0.53
Mean SES					
Random part					
Intercept school level	1.65	0.21	1.63	0.20	
Model fit					
Deviance	19350.7		19256.8		
AIC	19354.7		19264.8		

1) θ, the threshold (intercept) value is quite a low negative value (-2.24) → meaning that we have a low baseline prob. of repeating the class.
→ In this case, baseline prob. corresponds to females who do not have pre-primary education (PPED).
& just by looking at this threshold value w/o doing any calculations, becuz it's -2 ..., we already know that at least less than half of the people, looking at the baseline prob. of repeating a class is less than 0.5.

Thailand Education Data

Model	Empty model		Predictors level 1, fixed effects		
Predictor	par est	SE	par est	SE	Odds ratio
Fixed part					
threshold	-2.23	0.08	-2.24	0.10	
Sex (female=ref)			0.54	0.07	1.71
PPED (no=ref)			-0.64	0.10	0.53
Mean SES					
Random part					
Intercept school level	1.65	0.21	1.63	0.20	
Model fit					
Deviance	19350.7		19256.8		
AIC	19354.7		19264.8		

1) DIY: what can we say in general on the probability of repeating a class?

2) DIY: what is the predicted probability of repeating a class for males? And for females?
(Given no pre-primary education)

3) How does this fit with the odds ratio?

2) prob. female: $\frac{e^{-2.24}}{1+e^{-2.24}} = 0.096$

prob. male: $\frac{e^{-2.24+0.54}}{1+e^{-2.24+0.54}} = 0.154$

This is how you can calculate the predicted probabilities for certain ppl in your class.

3) odds ratio = $\frac{\frac{0.15}{1-0.15} \text{ odds for male}}{\frac{0.096}{1-0.096} \text{ odds for female}} = 1.71$

|| ratio of ratio || $\left(\frac{\text{odds of repeating a class for males}}{\text{odds of repeating a class for females}} \right)$

you can also see again this is what odds ratio relates to: So how does the odd changes given that you're male instead of female.

3/3/22

Again: definitions of logit and odds

p_{ij}=probability of repeating a class ~ which is an abstract latent construct.

neg. effect of PPED: children w/ pped are less likely to repeat a class

$$\text{logit}(p_{ij}) = 2.237 + 0.536sex_{ij} - 0.642pped_{ij} + u_{0j}$$

threshold

pos effect of sex: boys are more likely to repeat a class

$$odds = \frac{p_{ij}}{1 - p_{ij}}$$: prob. of success / prob. of failure

β coefficient we had in our model

$$odds\ ratio = \frac{odds_{males}}{odds_{females}} = e^{0.536} = 1.709$$ obtained by taking exponent on the β coefficient value.

Interpretation of model parameters

Qualitative interpretation for sex

- Males have higher probability of repeating a class than females ∵ reg. coefficient is positive.

Quantitative interpretation for sex

- On logit scale: ⤳ log (odds)
  - Logit probability of repeating a class is 0.536 higher for males than for females (effect is additive on logit scale) : but it doesn't really say much, cuz it's wired S-shaped curve thingy....
- Using odds ratios
  - Odds of repeating a class are 1.709 times as high for males as for females (effect is multiplicative on odds scale) odds for male is 70% higher than females. (above 1)

what if 0.709 (below 1) : you have to in your mind 1 − (sth) then you know how much percent it decreases. So here, 30% less

Thailand Education Data

- $\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2+\sigma_R^2} = \frac{1.725}{1.725+3.29} = 0.331$

- Estimates of variance components
  - $\sigma_{u0}^2 = 1.627$
  - $\sigma_R^2 = 3.290$
  - $\sigma_F^2 = 0.176$ var. of linear predictor
- Explained and unexplained variance
  - Explained: $\frac{\sigma_F^2}{(\sigma_F^2+\sigma_{u0}^2+\sigma_R^2)} = 0.035$ calculate total explained var. which is only about 4%
  - Unexplained level 1: $\frac{\sigma_R^2}{(\sigma_F^2+\sigma_{u0}^2+\sigma_R^2)} = 0.646$
  - Unexplained level 2: $\frac{\sigma_{u0}^2}{(\sigma_F^2+\sigma_{u0}^2+\sigma_R^2)} = 0.319$

Explained part is given by var. of linear predictor

2 unexplained var. components

We can ofc recalculate these, when we add or leave out certain predictors.

Emmeke Aarts　　Multilevel analysis - lecture 3a　　39

39

Multilevel ordinal logistic model

Ordinal outcomes

Emmeke Aarts　　Multilevel analysis - lecture 3a　　40

40

Ordinal Data

- Typical for Likert items such as:
  1. decidedly *no*
  2. more *no* than *yes*
  3. more *yes* than *no*
  4. decidedly *yes*
- Or frequencies of behavior such as: ... occurs
  1. never
  2. sometimes
  3. always

We're still doing logistic regression, we're still treating them as YES/NO outcomes.
But here, we're "combining categories"

The Cumulative Regression Model

- Assign simple consecutive values to the *C* ordered categories, such as 1...*C*
- For a response variable *Y* with 3 categories ('never', 'sometimes' & 'always') we have three response probabilities *p* plus cumulative response probabilities *p**

$$\text{Prob}(Y=1) = p_1 \qquad p_1^* = p_1$$ 'never'

$$\text{Prob}(Y=2) = p_2 \qquad p_2^* = p_1 + p_2$$ 'never' or 'sometimes'

$$\text{Prob}(Y=3) = p_3 \qquad p_3^* = p_1 + p_2 + p_3 = 1$$ ← redundant all possible outcomes

Two-level Ordinal Response Model

- p_1^* and p_2^* are cumulative probabilities becuz p_3^* is redundant.
- generalized linear mixed model can be used to analyze the cumulative probabilities → this is what we're doing w/ ordinal regression

set of

- Logistic regression: $logit(p_c^*) = log\left(\frac{p_c^*}{1-p_c^*}\right) = \eta_c$

they get their own intercept

for each cumulative outcome

- Note that a different intercept is specified for each p^*, the thresholds θ_c

besides that all other reg. coefficients are same over the diff. outcome classes!

Emmeke Aarts Multilevel analysis - lecture 3a 43

43

Latent variable presentation

- Assume a continuous latent variable underlying the categorical responses
- The observed categorical response is related to the position on the latent variable by a threshold model

$$y_{ij} = \begin{cases} 1 & if \quad -\infty < \eta_{ij} \leq \theta_1 \\ 2 & if \quad \theta_1 < \eta_{ij} \leq \theta_2 \\ \dots & \dots \\ C & if \quad \theta_{C-1} < \eta_{ij} < \infty \end{cases}$$

: If you're below the first threshold θ_1, you're in the outcome category #1

: If you're between θ_1 & θ_2, you're in outcome category #2.

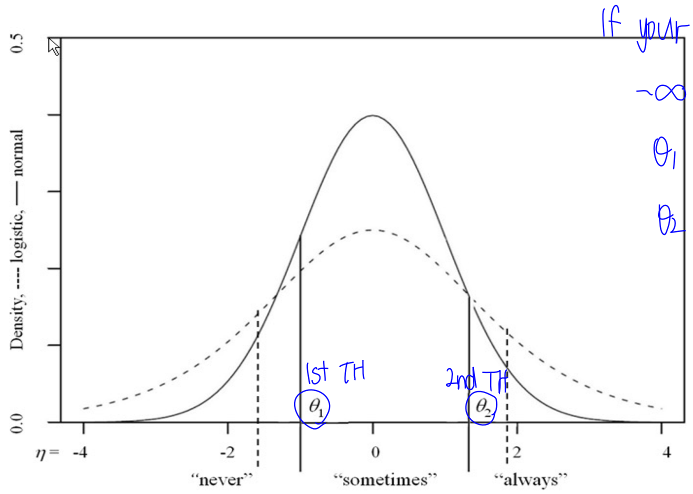
⋮

: If you're above θ_{C-1}, the # of category −1th threshold, then you're in the last outcome category, C.

Emmeke Aarts Multilevel analysis - lecture 3a 44

44

Latent variable presentation



If your predicted outcome on η scale is
$-\infty < \quad < \theta_1$: "never" category,
$\theta_1 < \quad < \theta_2$: "sometimes" category
$\theta_2 < \quad < \infty$: "always" category

45

Now if we make a separate reg. equation for each of the cumulative outcome category:
η_{1ij} → "never" outcome category
η_{2ij} → "never" + "sometimes" outcome category

Two-level Ordinal Response Model

- The single equation version is:

only thing that changes for this second cumulative probability is that threshold θ_2! All the other things are identical, they're fixed over diff. models!

$$\eta_{1ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}$$
$$\eta_{2ij} = \theta_2 + \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}$$
$$\dots$$
$$\eta_{Cij} = \theta_C + \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}$$

so we can also put them in a more compact way :)

- Which can be written in a much compacter way as:

$$\eta_{Cij} = \boxed{\theta_C} + \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}X_{ij}Z_j + u_{0j} + u_{1j}X_{ij}$$

with $\theta_1 = 0$ threshold specific to the cumulative category (for the first outcome category, this would be zero)

"Proportional odds assumption" All the other things are fixed over diff. models!

the effect of reg. coefficients on probability of being in a specific cumulative class. those effects are assumed to be fixed.

46

the effect of predictors are fixed across diff. categories!

An example

Of logistic multilevel regression on ordinal outcomes

Feeling unsafe on the street

- Survey to assess how characteristics of streets affect feelings of unsafety in people walking these streets
- A sample of 100 streets is selected, and on each street a random sample of 10 persons is asked how often they feel unsafe while walking that street
- The unsafety is asked using three answer categories
  - 1 = never, 2 = sometimes, 3 = often

Safety data

- Predictors:

level 1 predictors (
  - age
  - gender (0=male, 1=female)

level 2 predictors (street level)
  - street characteristics
    - economic index (standardized *Z*-score)
    - crowdedness rating (7-point scale)
- All predictors are centered on the grand mean, and age is divided by 10

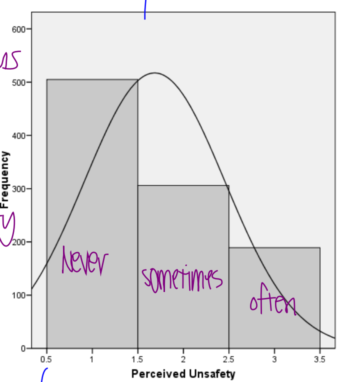
is done so that the range of the predictors is not so hugely divergent.

Safety data

see about 50% of ppl that we asked "never" feel unsafe & about 30% "sometimes & about 20% "never" feel unsafe.

- The perceived unsafety is asked using three categories
  - 1 = never, 2 = sometimes, 3 = often
- Data are ordinal categorical, useful when instead of continuous interpretation,,
  - Few categories
  - Skewed distribution ~ if we don't really have normal dist. of outcome, we already know that we are quite likely to violate the normality assumption of residuals.

w/ordinal model, we use
- Cumulative probabilities:
  - $p_1^* = P(never)$
  - $p_2^* = P(never) + P(sometimes)$
- For both a binary logistic regression model is fitted



If we don't use ordinal reg, but use normal regression, we'd run into problems, becuz this is not normal at all ⇒ super skewed! not symmetric.

- General intercept: relates to the $p_1^* \sim$ cumulative prob. of "never" feeling unsafe
0.03 = almost zero, which means that for males (reference category) about 50% of males in our data set never feel unsafe (predictors are centered)
relate that to the histogram above, this makes sense! bcuz this is about 50% of the data.

3/3/22

- If we look at the 2nd threshold, can we use that in anyways? For males, we see that the prob. of being in either "never" or "sometimes" is really high (1.91). And this is almost 2, so that means combined prob. of being in either of those classes is about 85%! If you want to know being in "sometimes" class, you subtract the $\theta_2 - \theta_1 = 35\%$, which means that prob. of being "sometimes" class is about 35%.

relates to the combined $p_2^* \sim$ either being in "never" or "sometimes" class

Safety data

Model	Fixed effects predictors levels 1 and 2		
predictor	par est	SE	OR
Fixed Part			
Intercept / threshold	0.03	0.09	
Threshold 2	1.91	0.11	
Age/10	0.41	0.05	1.51
Gender (ref=male)	1.13	0.14	3.09
Economic	-0.69	0.08	0.50
Crowded	-0.44	0.05	0.65
Random part			
Random intercept	0.20	0.10	
Model fit			
Deviance	1739.7		
AIC	1753.7		

Emmeke Aarts Multilevel analysis - lecture 3a

- AGE. coef is positive: 0.41. not very big though. As you grow older, you're more likely to instead of being in "never" category, to be in "never" & "sometimes" category & if you grow older, you're more likely to instead of being in "never + sometimes" category, be in the "always" category ⇒ So you're more likely to move one category up! & how likely you're to move one category up? It doesn't matter. If you're switching from never → sometimes or sometimes → always that proportion is the same!

51

- If you're female instead of male, you're also more likely to be in a higher category with respect to how often you feel unsafe.

- If the street has a higer economic score, ppl are less likely to be in a higher unsafe category (−0.69), If it's crowded, also less likely to be in a higher unsafe category

Interpretation of model parameters

Qualitative interpretation for sex

- Females are more likely to feel unsafe more often

Quantative interpretation for sex

- On logit scale
  - Logit probability of (sometimes, often) versus (never) is 1.13 higher for females than for males
  - Logit probability of (often) versus (never, sometimes) is 1.13 higher for females than for males
- Using odds ratios
  - Odds of (sometimes, often) versus (never) 3.09 times as high for females as for males
  - Odds of (often) versus (never, sometimes) 3.09 times as high for females as for males

Emmeke Aarts Multilevel analysis - lecture 3a 52

52

This two things, they're the same → $\boxed{\text{proportional odds assumption}}$, maybe it makes sense here but sometimes it's very strange thing to your data.

*Take home Msg: The effect of sex on moving a category up, that's the same, it doesn't matter which category you move up from!! across diff. outcome categories

Interpretation of model parameters

Proportional odds model: effects of predictors are equal for the regression models for p_1^* and p_2^*. This simplifies quantitative interpretation.

we can just say

Quantitative interpretation for sex

- On logit scale
  - Logit probability is 1.13 higher for females than for males
- Using odds ratios
  - Odds are 3.09 times as high for females as for males

53

Summary

- When analyzing dichotomous outcomes using (multilevel) logistic regression, we are modelling the log of the odds
- When analyzing ordinal outcomes using (multilevel) logistic regression, we are modelling the log of the *cumulative* odds ~so we're combining classes
- for both: The lowest level residual error is fixed for all models, to $\frac{\pi^2}{3} \approx 3.29$ for logistic regression
  so just passing on what's the effect on the outcome
- Regression coefficients β are interpreted qualitatively, the odds ratios (the exponent of β) are interpreted quantitatively. One can also obtain the predicted probability of belonging to a certain category given the predictors
- To calculate the explained and unexplained variance at each level, we need adjusted equations that includes the fixed variance σ_F^2

fixed var. of linear predictor becuz we don't have the residual var. at the lowest level (it's fixed)

54

Final comments

- Estimation in multilevel logistic regression models is difficult
  - Compare output of different software packages (or estimation methods)
  - Use different starting values and compare output
  - Sample size should be sufficiently large
- When to treat ordinal data as continuous?
  - With ≥7 ordered categories → but then all the categories needed to be filled. Often, ppl only use 3-6 for ex & that doesn't count as 7 categories. It's about how many of them are used.
  - With 5 or 6 categories IF all categories used and more or less normal distribution
    really the minimum
    there has to be a peak which is surrounded by something that goes down on both sides.. So it can be unsymmetrical but it has to be reasonable...
- Also generalized linear mixed models for
  - Nominal outcomes
  - Count outcome

55

Questions?

56