

Feb 21

- 1. Final project topic info**
- 2. Cumulative probability distributions**
- 3. Predicting educational attainment**
- 4. Log cumulative odds link**
- 5. The ordered logit model**
- 6. Estimating ordered logit in R**

See also: Michael Betancourt's more technical introduction to ordered logit:

https://betanalpha.github.io/assets/case_studies/ordinal_regression.html

Final project topics

Final projects should:

- ∴ Use at least one of the ‘advanced’ methods from the course (e.g. logit, multinomial, multilevel, missing data imputation, ...)
- ∴ Thoroughly motivate, estimate, and interpret the results from at least one such model

Tips on data sources (not required, but helpful)

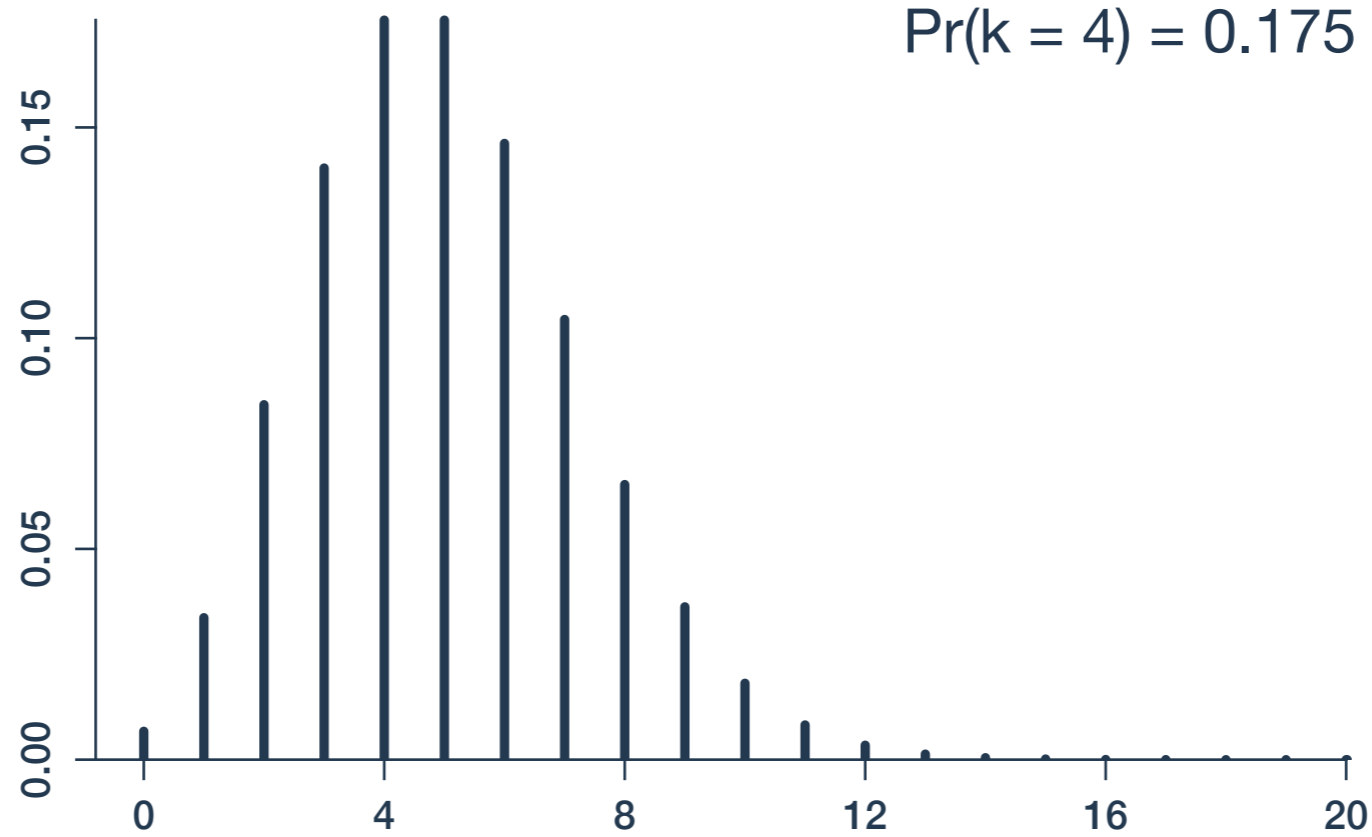
- ∴ Pick something you find interesting and useful!
- ∴ Look for something with either a non-continuous outcome variable (for generalized linear models) or meaningful clustering of observations (for multilevel models).
- ∴ Minimize the need for data merging and cleaning — try to find something that you can download as a nice single table
- ∴ Don’t worry too much about the data format — R (with the ‘rio’ package) can read data from SPSS, SAS, Stata, Excel, etc.

Think of this as an opportunity

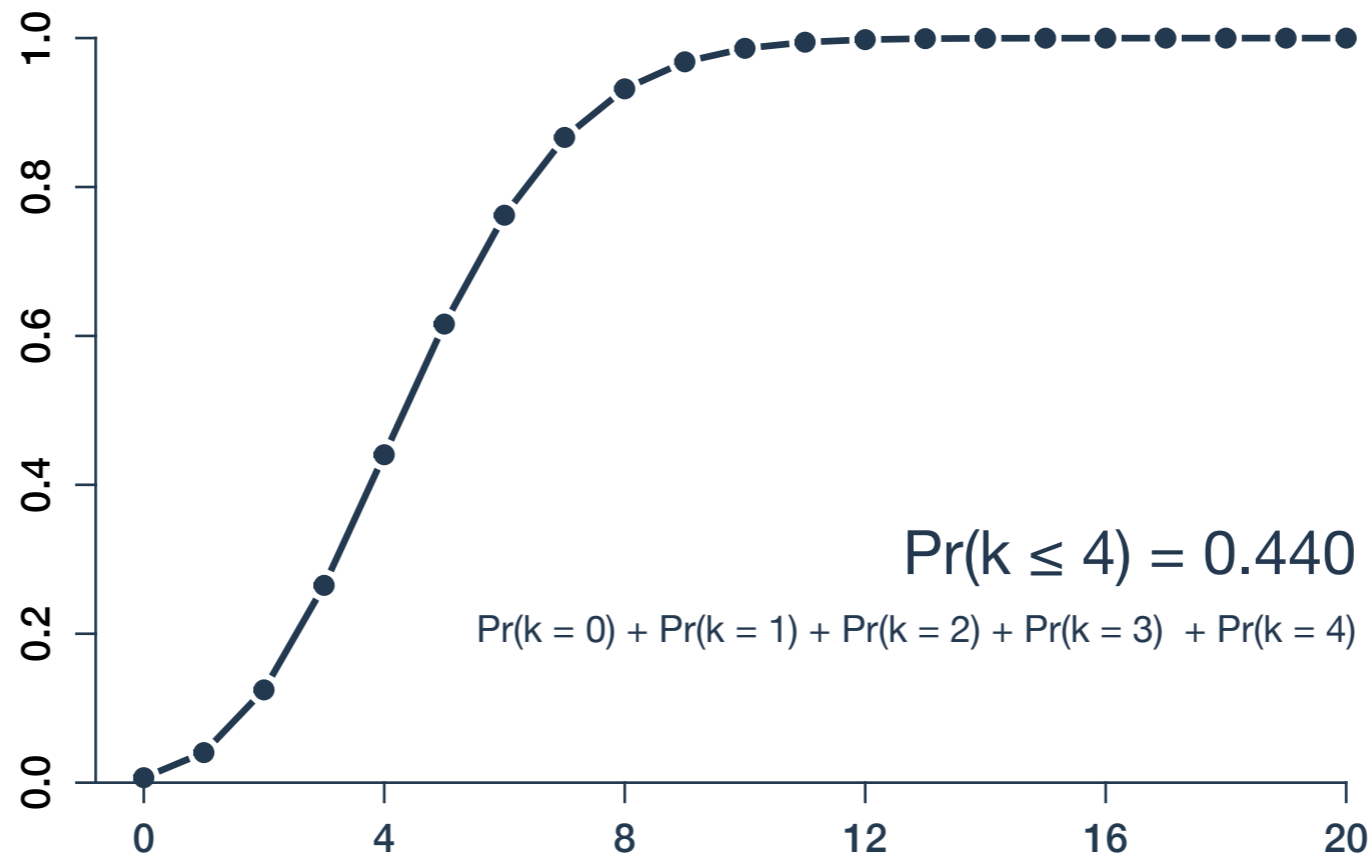
- ∴ Could this project be a pilot for a dissertation chapter? Could it dovetail with the paper for another class you’re taking? Could it let you try out a wild idea for your qual data?

Cumulative probability distributions

Probability mass function (PMF)
Pois(5)



Cumulative distribution function (CDF)
Pois(5)

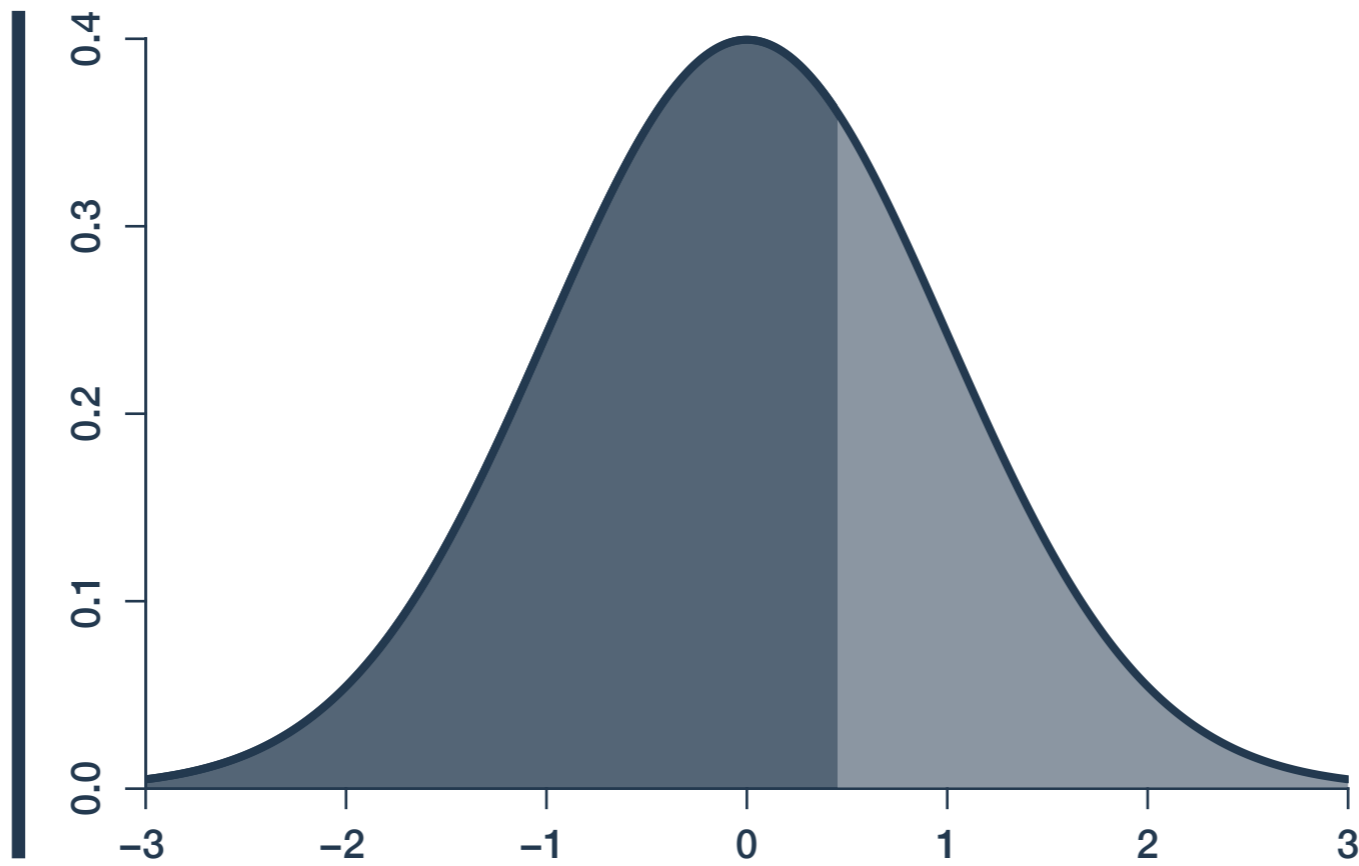


Two ways to describe the exact same abstract Poisson distribution

Cumulative probability distributions

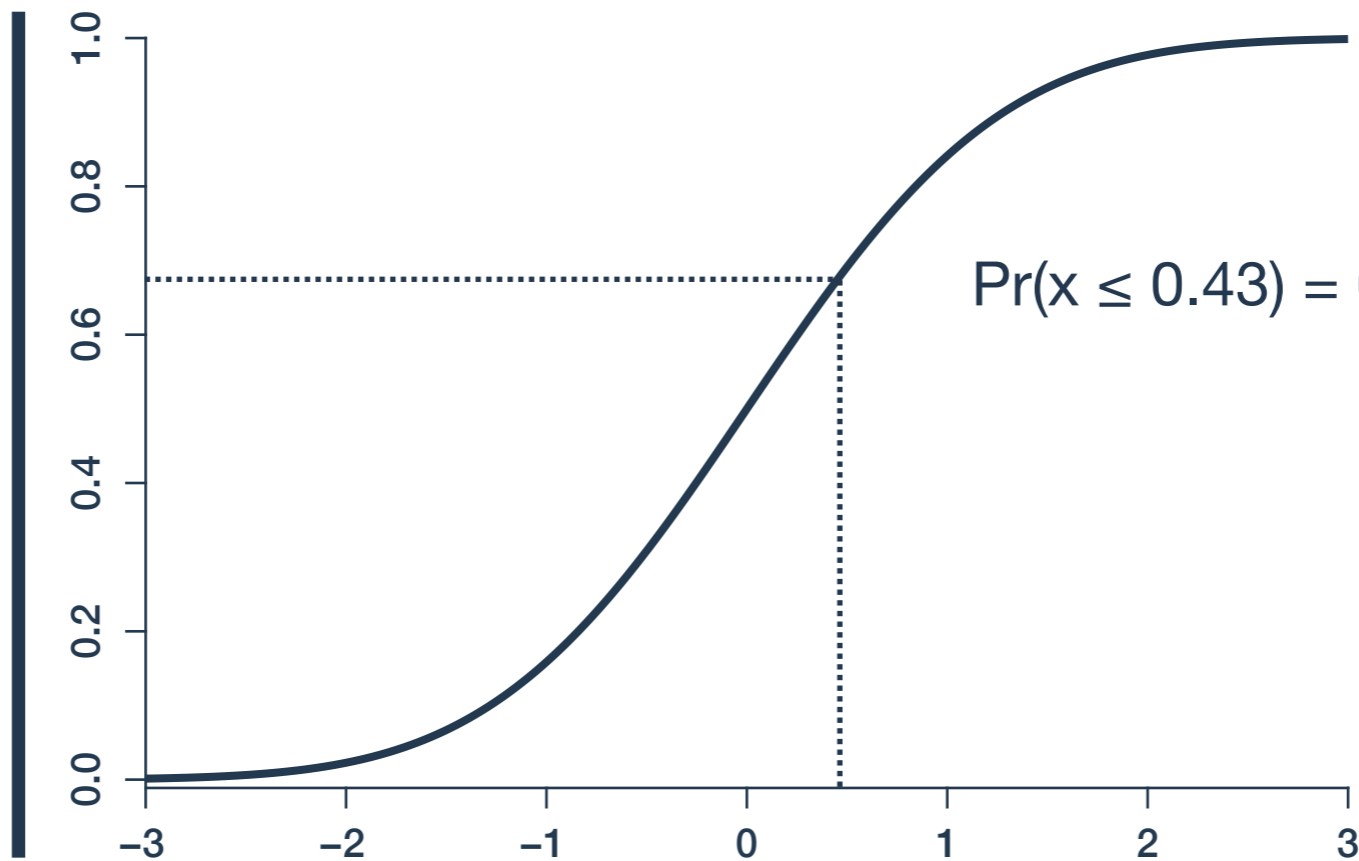
Probability density function (PDF)

Norm(0, 1)



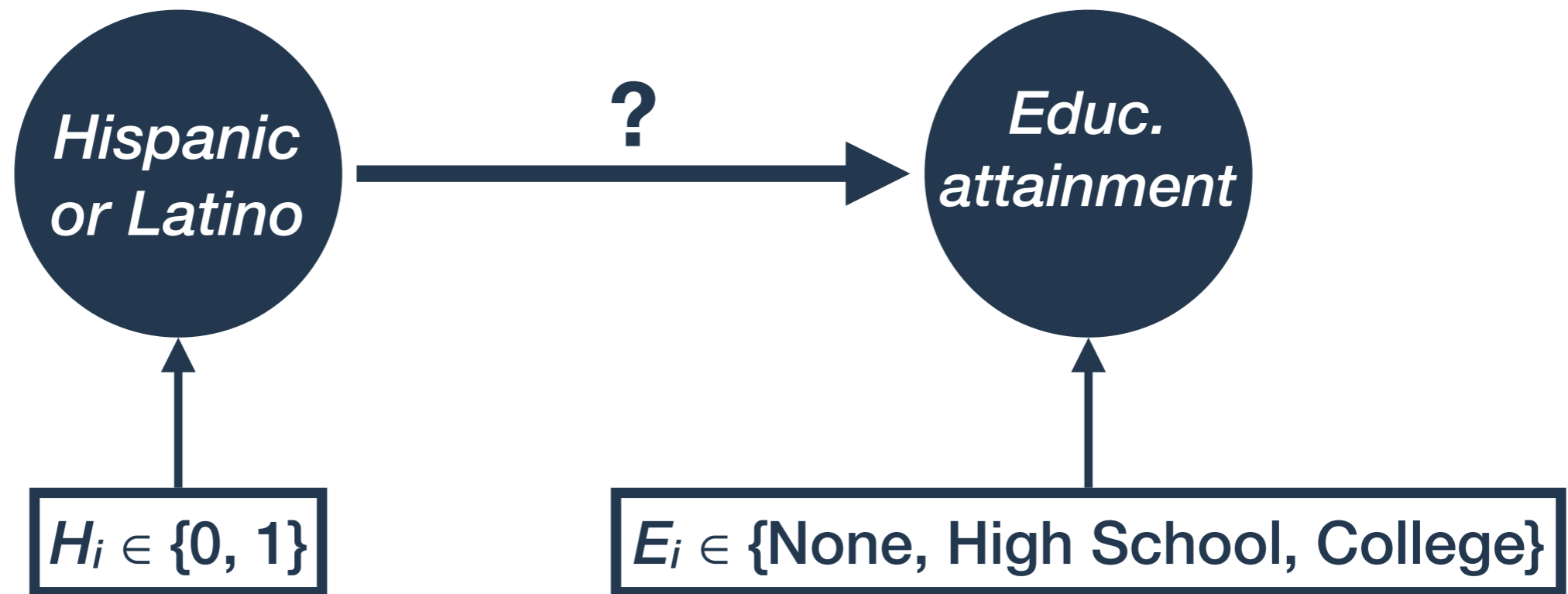
Cumulative distribution function (CDF)

Norm(0, 1)



Two ways to describe the exact same abstract Normal distribution

Age and education



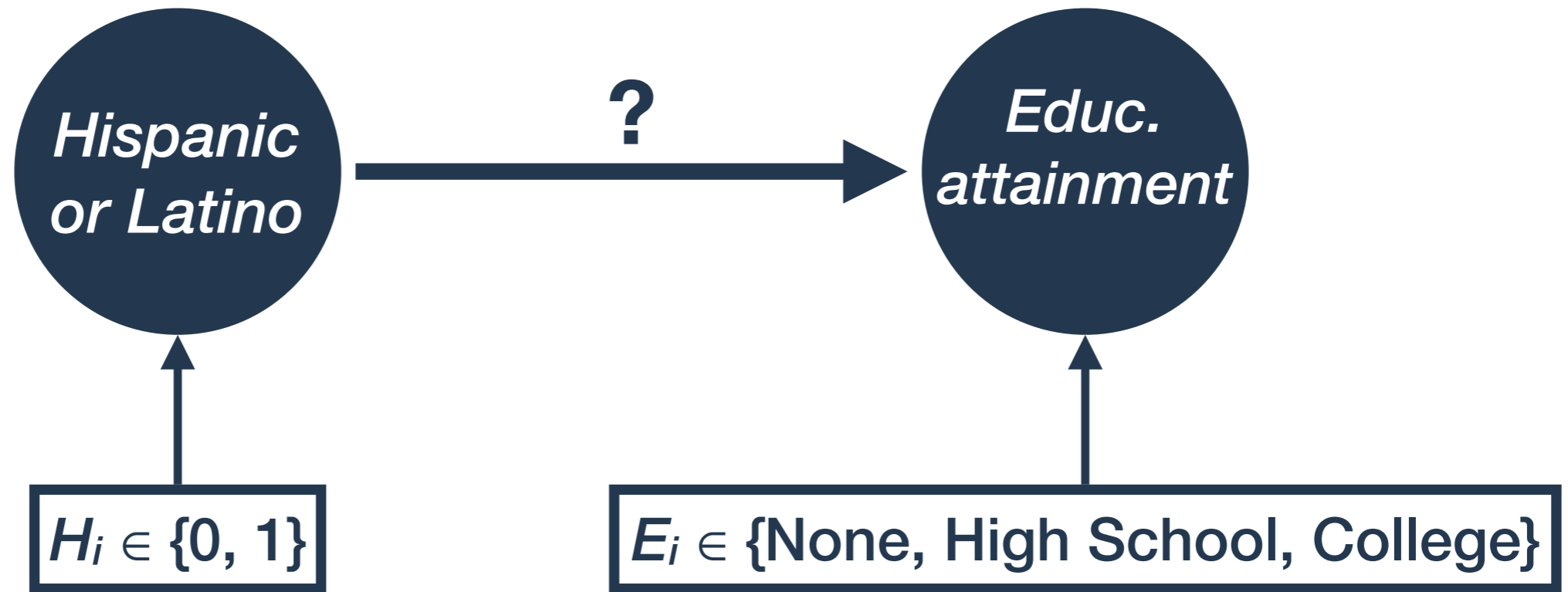
The problem

Educational attainment is measured as categories, but those categories are ordered in an important way.

The solution

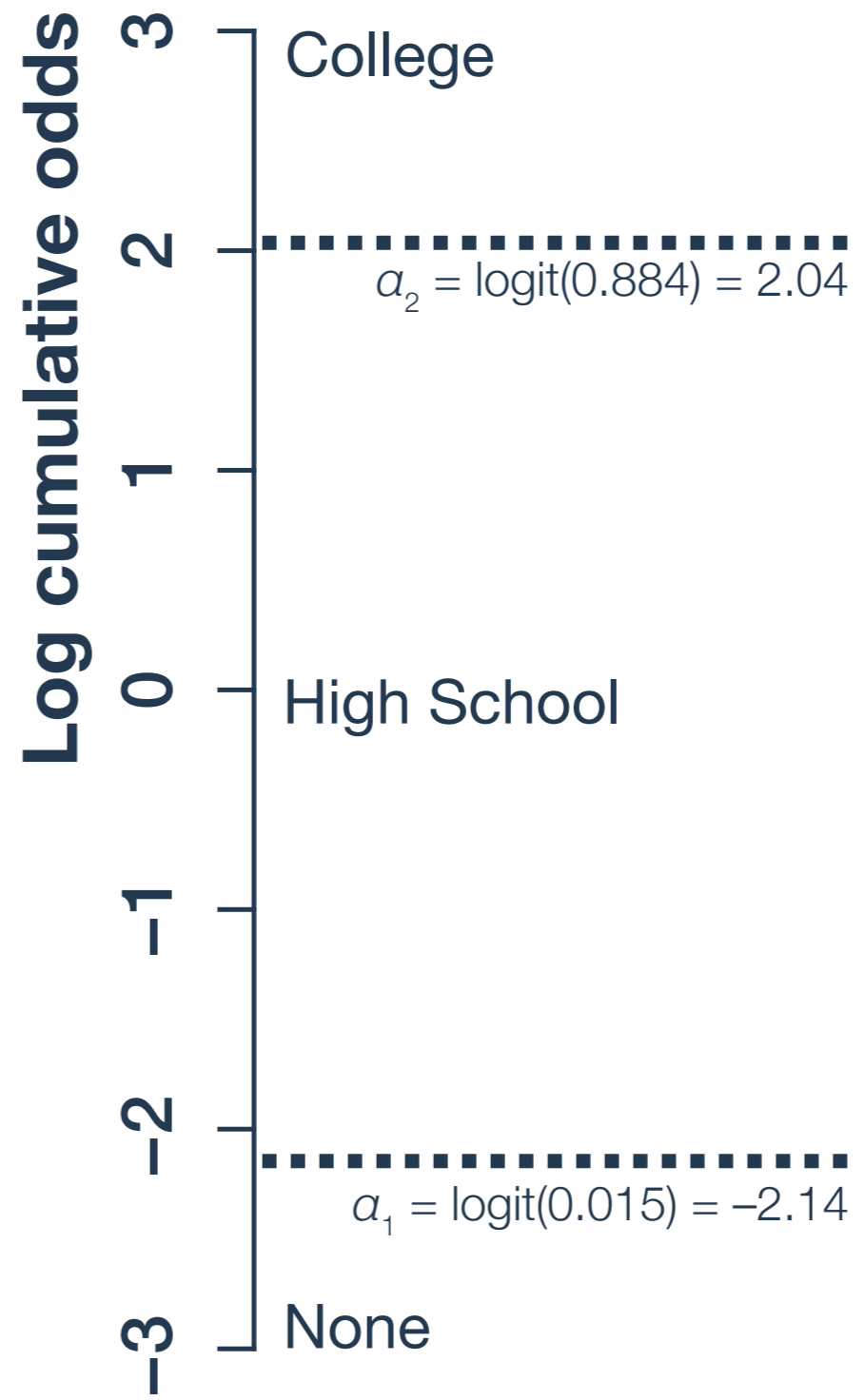
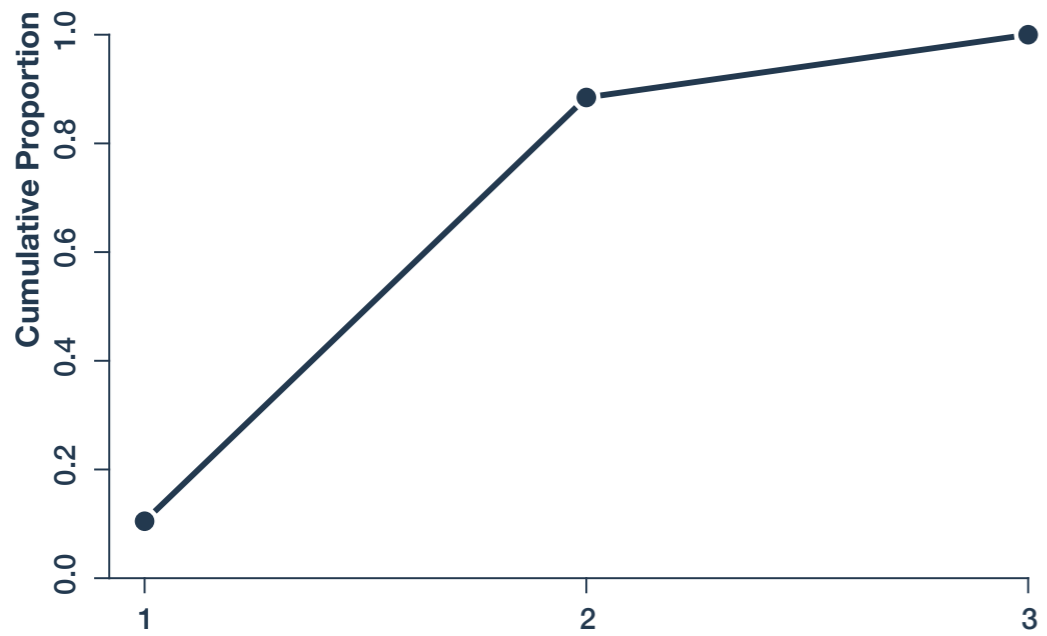
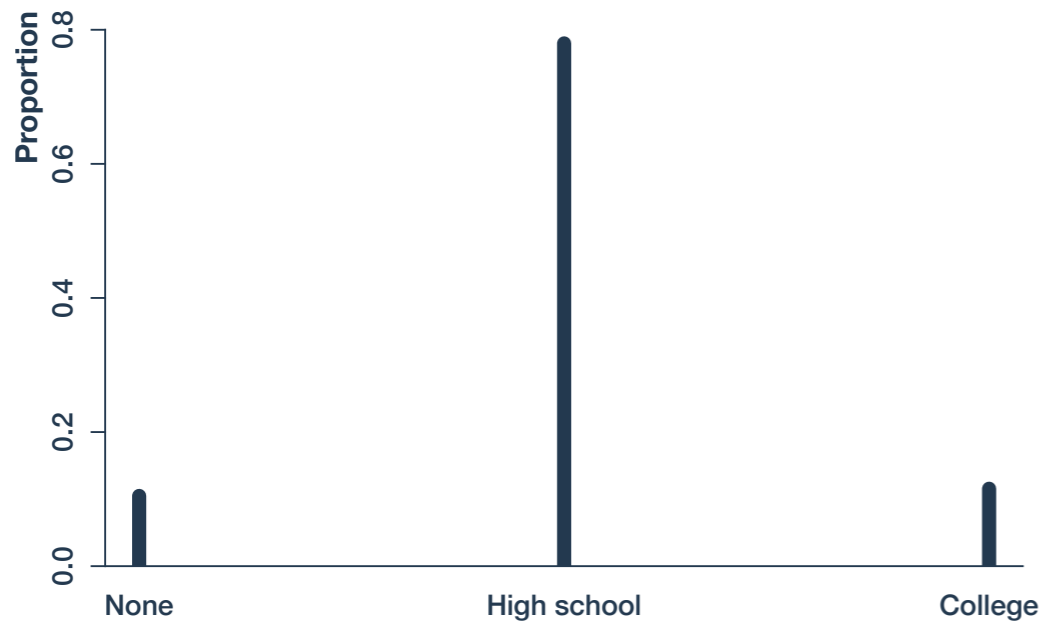
Treat education as a categorically-distributed variable, but constrain the probabilities to respect the order.

Age and education



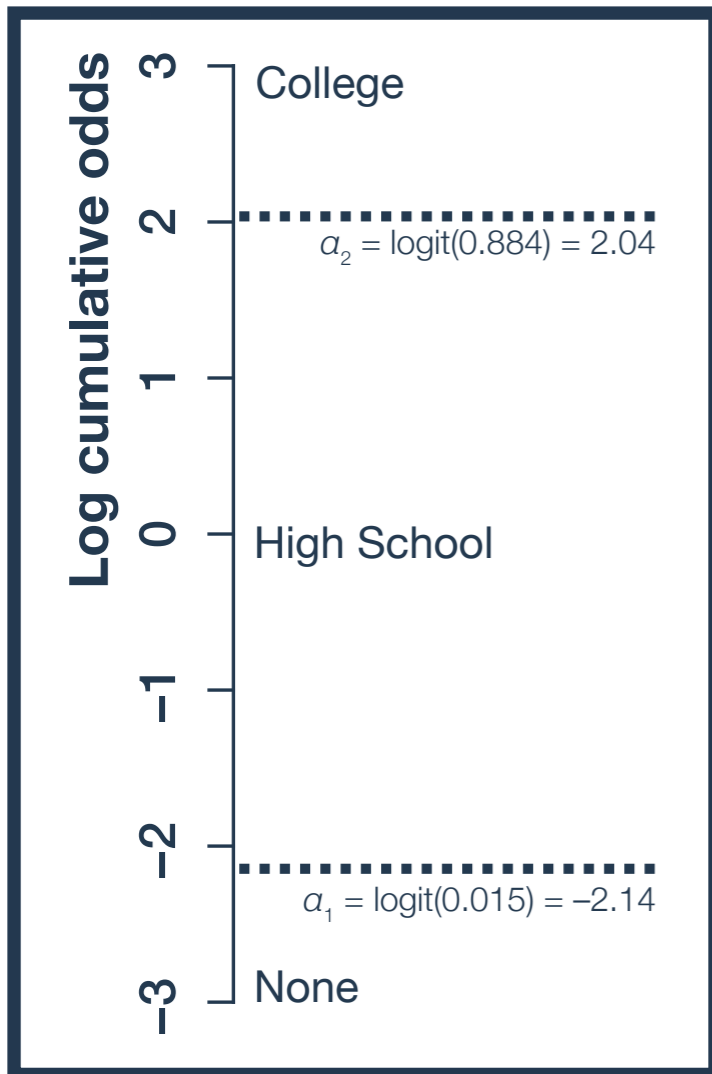
$$E_i \sim \text{Categorical}(p_1, p_2, p_3)$$

Log cumulative odds



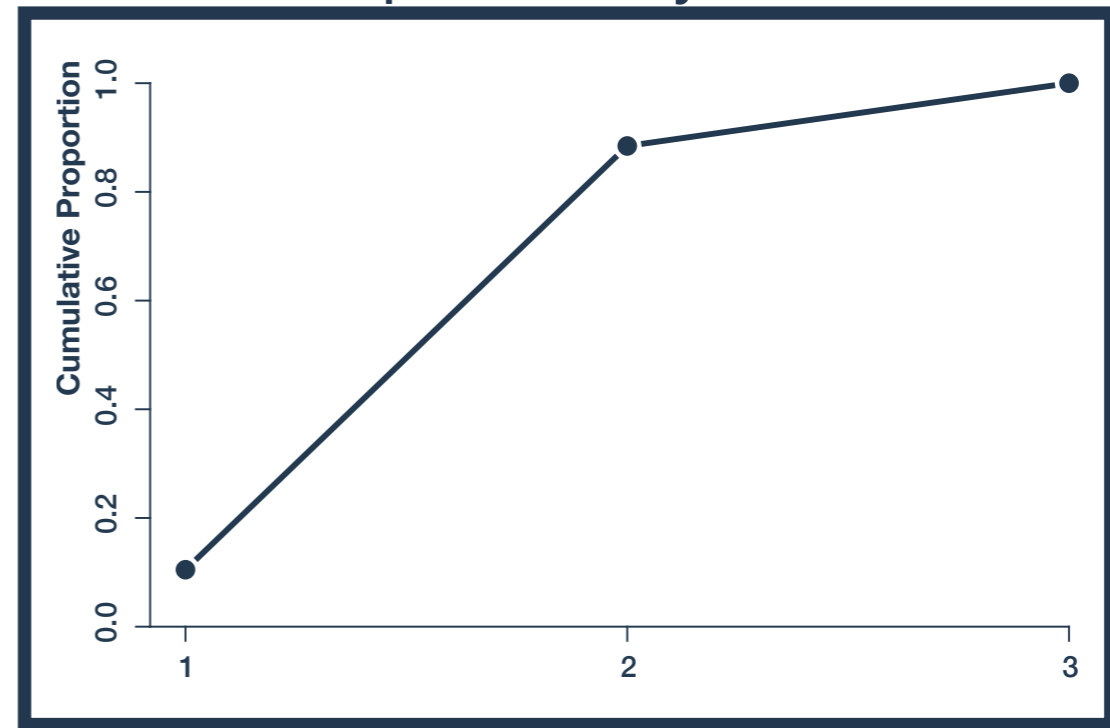
Log cumulative odds

Parameter scale



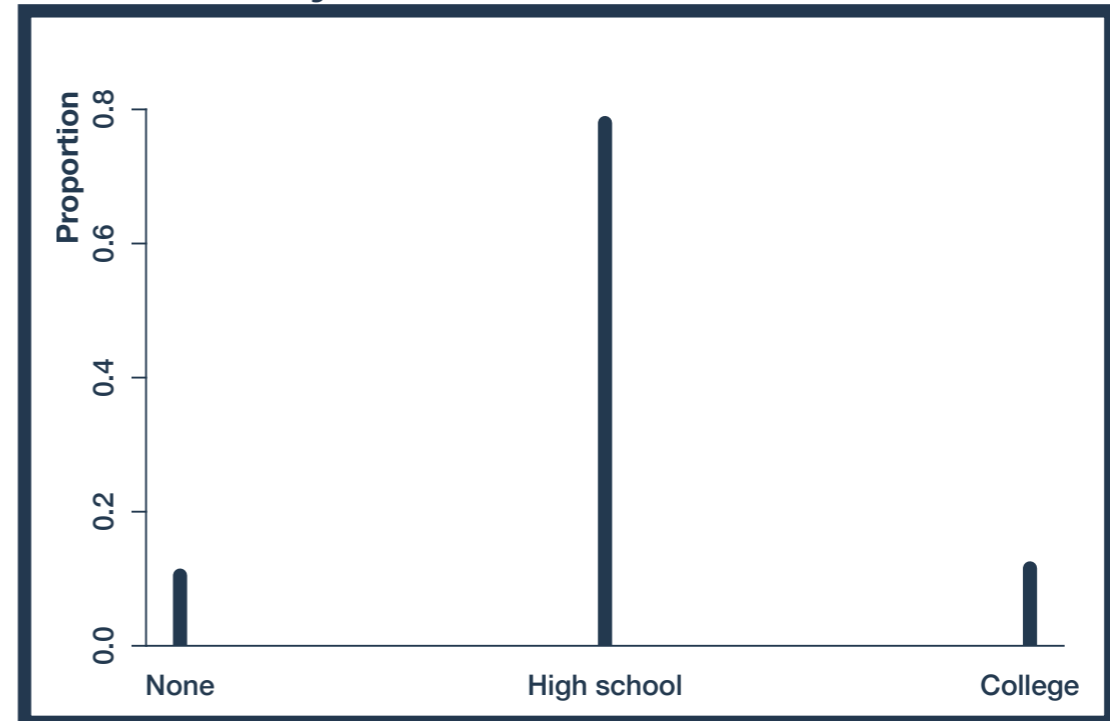
step 1
Inverse logit

Cumulative probability



step 2
Subtraction

Probability



**A two-step
link function**

Ordered logit model

Intercept-only
ordered logit
model

$$E_i \sim \text{Categorical}(\mathbf{p})$$

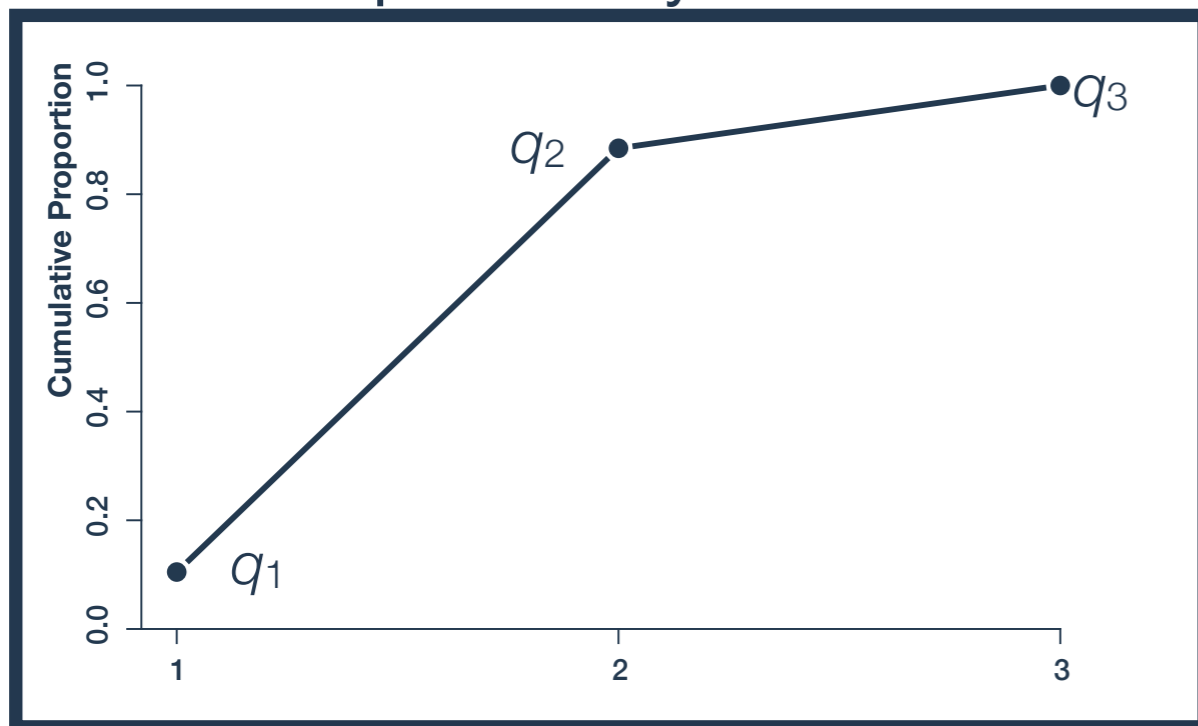
$$p_k = q_k - q_{k-1}$$

$$\text{logit}(q_k) = a_k$$

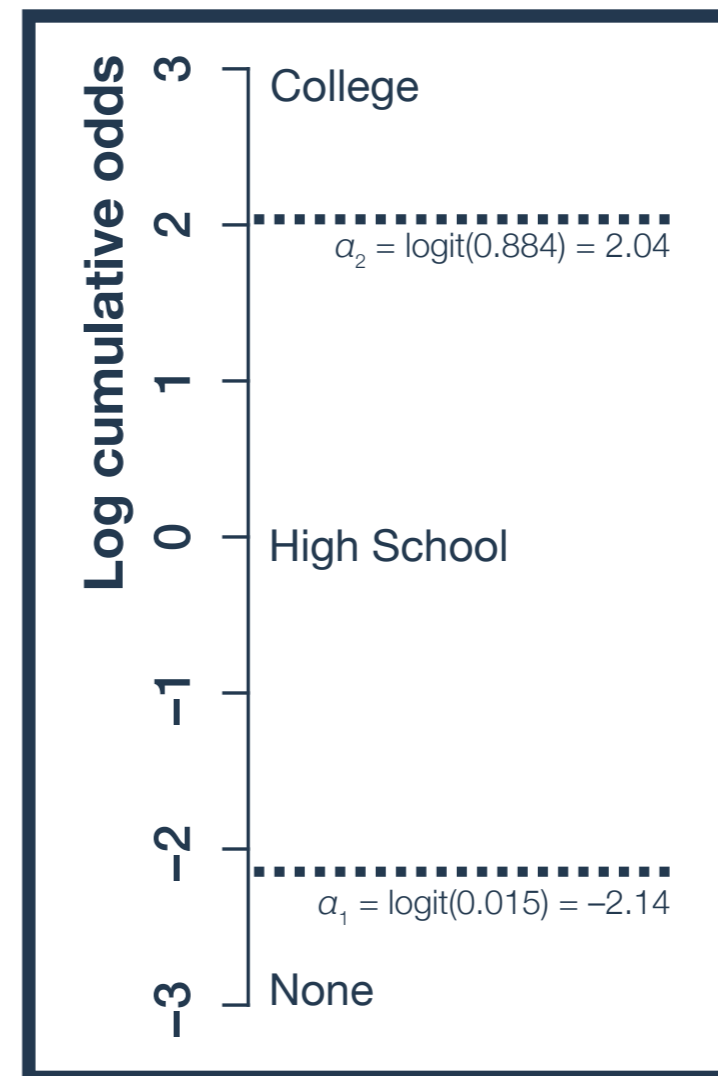
q_k is the cumulative
probability of category k

$$q_0 = 0; q_k = 1$$

Cumulative probability



Parameter scale



Ordered logit model

Ordered logit with predictors

$$E_i \sim \text{Categorical}(\mathbf{p})$$

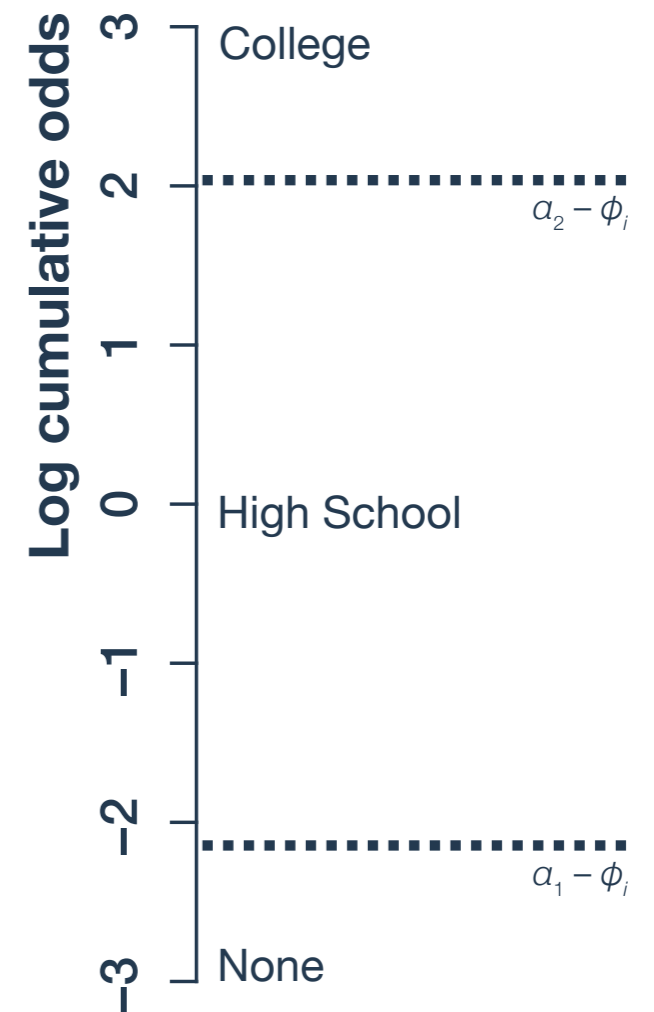
$$p_k = q_k - q_{k-1}$$

$$\text{logit}(q_k) = a_k - \phi_i$$

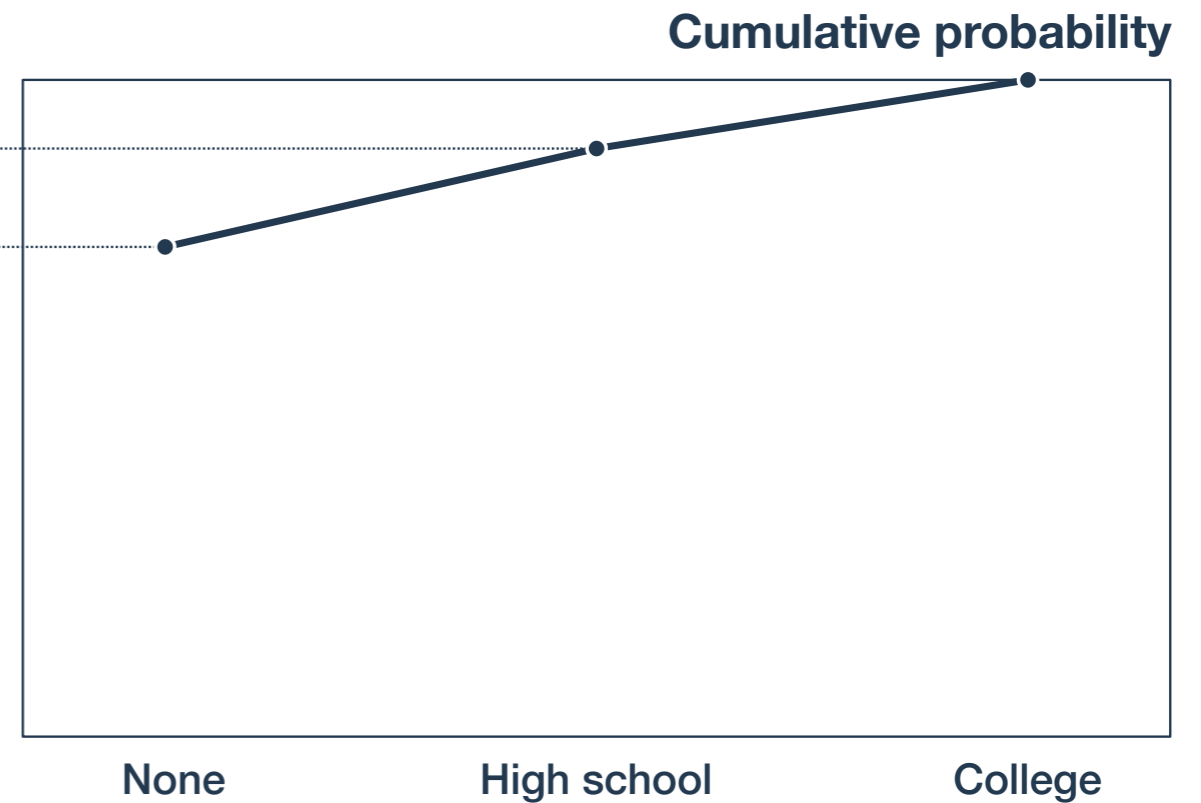
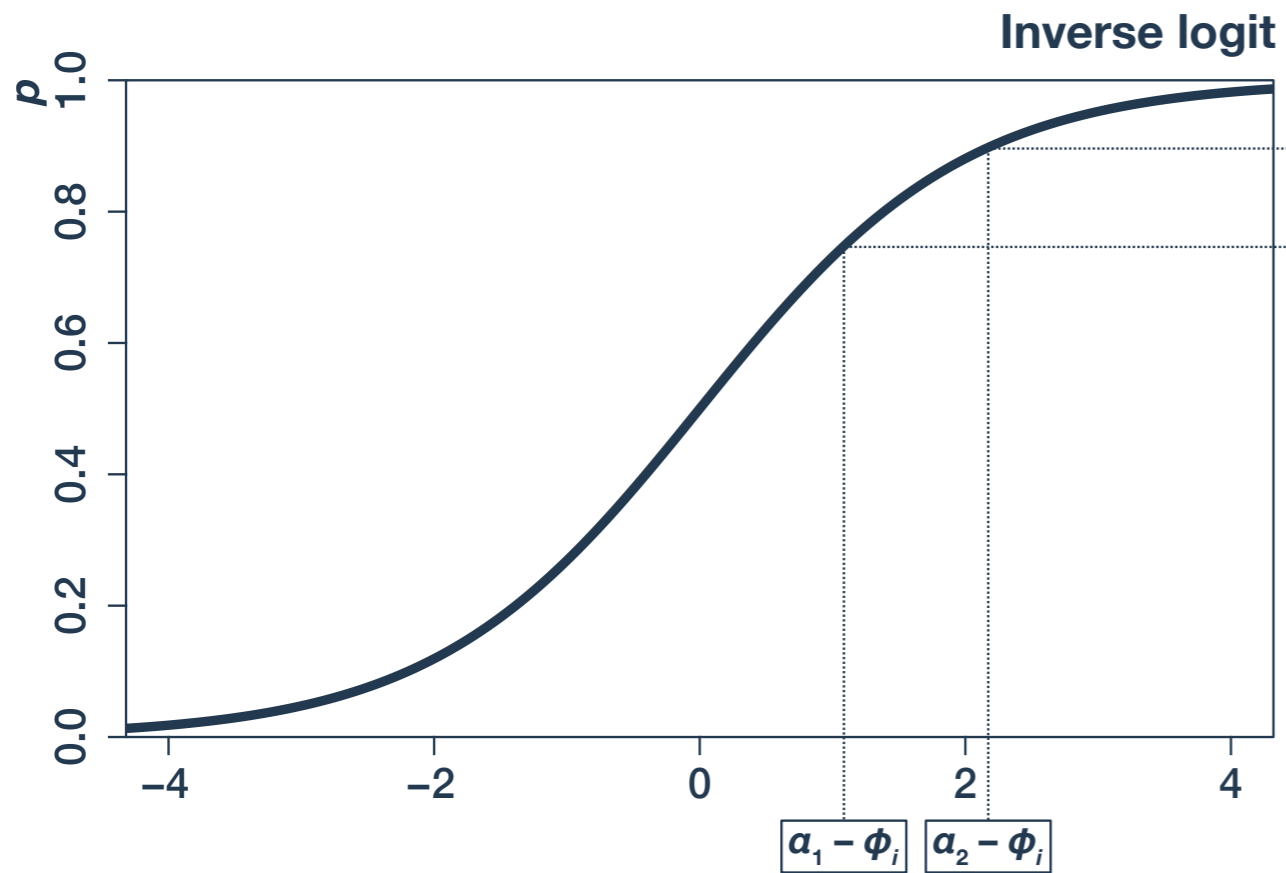
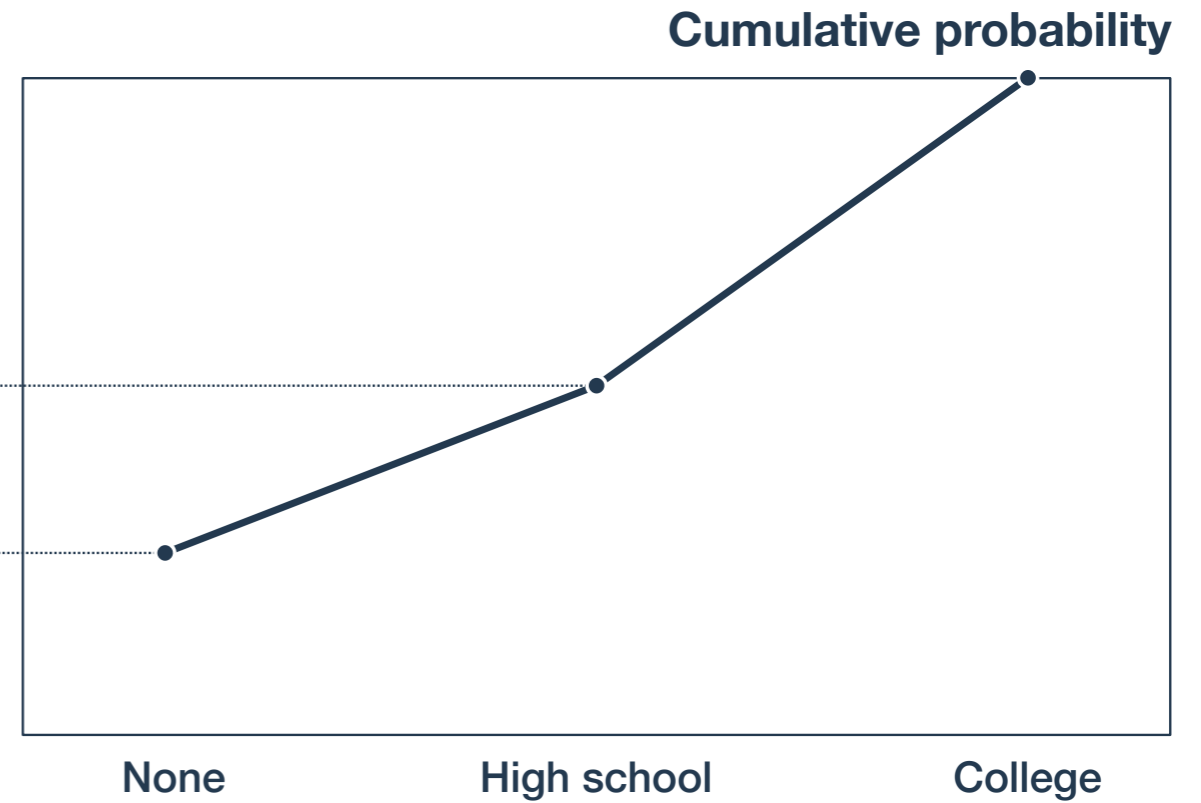
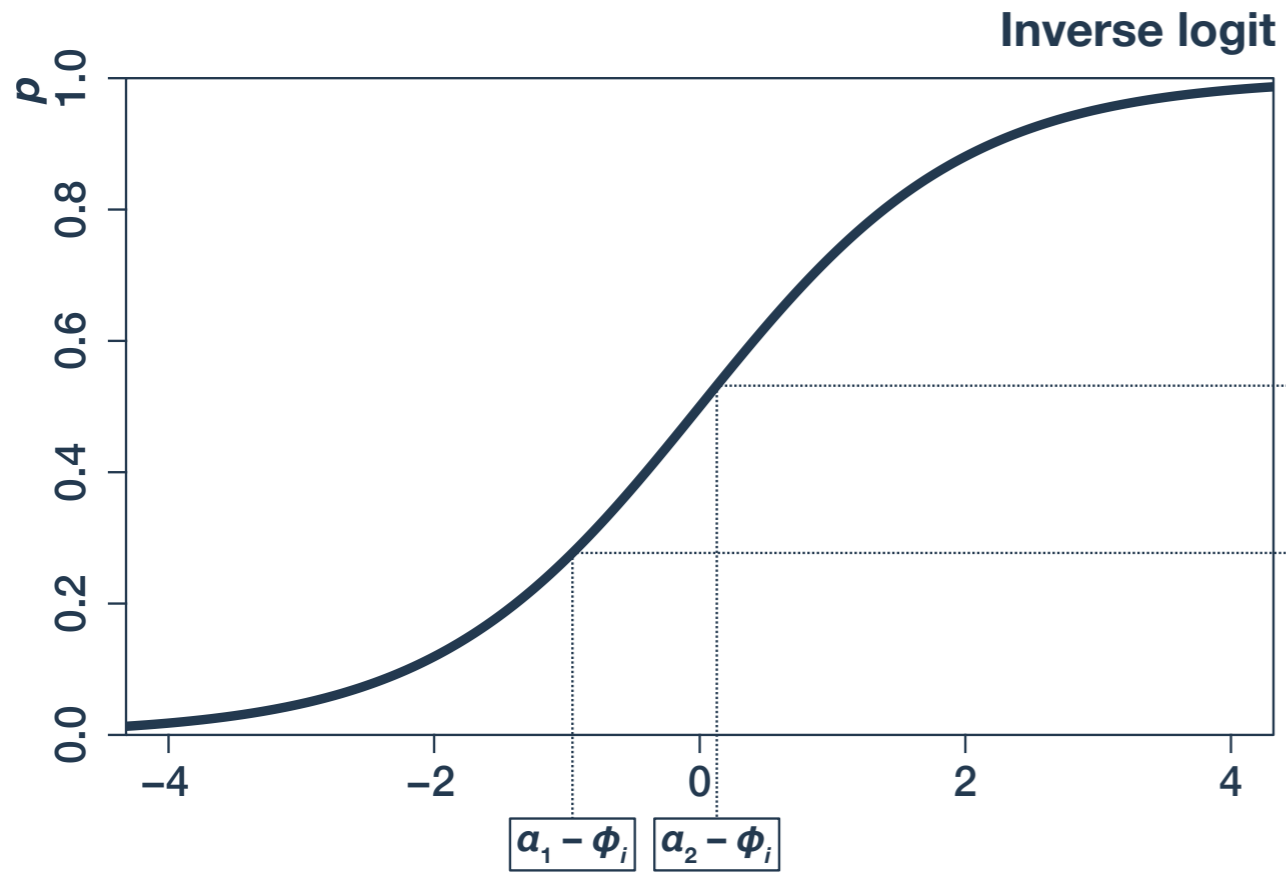
$$\phi_i = \beta H_i$$

$$a_k \sim \text{Norm}(0, 1.5)$$

$$\beta \sim \text{Norm}(0, 2)$$



Ordered logit model



Ordered logit model

