Models of outcome and choice: The logit model

Silje Synnøve Lyder Hermansen

November 14, 2019

Table of Contents

Section 1

A latent variable approach to GLMs

Many outcomes are not continuous

OLS assumes a continuous dependent variable. But many phenomena in the social sciences are not like that.

- Vote choice, civil conflict onset, legislator performance, court rulings, time to compliance, etc.
- What phenomena are you interested in?

$\Rightarrow$ *OK. Let's strategize.*

All regressions are linear(ized)

The basic formulation in any regression describes a linear relationship between x_i and y_i:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{1}$$

- When x_i increases with one unit, y_i increases with β units.
- If that relationship is not linear, we have to make it so:
  - by recoding the x_i
  - by recoding the y_i $\rightarrow$ we *linearize*.

A latent variable

A linear(ized) model requires a continuous dependent variable.

- Imagine we are interested in unobservable variable, z_i, that describes our propensity towards something.
- Above a certain threshold (τ) of z_i, observability kicks in and we can see y_i.
- The regression coefficients (β) in GLMs describe that relationship.

$\Rightarrow$ The latent variable approach is useful when interpreting the results.

Example: The binomial model

The logit model is a perfect example:

$$y_i = \begin{cases} 1 & \Leftrightarrow & z_i > \tau \\ 0 & \Leftrightarrow & z_i \leqslant \tau \end{cases} \tag{2}$$

- The probability (z_i) of an outcome y_i is continuous.
- Above a certain probability (τ), we observe a positive outcome ($y_i = 1$).

$\Rightarrow$ *but how do we set the value of τ?*

From latent variable to descrete outcomes

Statistical theory helps us describe how z_i leads to y_i.

- What kind of process generated our data? $\rightarrow$ data generating process (DGP)
- How can we best describe it? $\rightarrow$ choice of *probability distribution* (in GLM)

The three components of GLMs

When fitting the model, we need to make three choices:

- A linear predictor: βx_i.
- A probability distribution: they're all in the exponential family
- A recoding strategy

The three components of GLMs

In R this translates to two additional arguments compared to your usual OLS.

- A linear predictor: $\rightarrow$ (`y ~ x`).
- A probability distribution: $\rightarrow$ (`family =`)
- A recoding strategy $\rightarrow$ (`link = `).

Latent variable approach for interpretation

- The latent variable approach is useful when interpreting results.
- That's when we map *from* the latent variable *to* the observed outcome.

$\Rightarrow$ *When estimating the model, we have to go the other way 'round.*

Table of Contents

Section 2

Recoding: How do we get from a binary to a continuous variable?

Data structure

We can only observe the outcome produced by the latent variable. There are two data structures for binary data:

- classes of observations: e.g.: rats in a cage, coin tosses...
- case-based: e.g.: legislator votes, Brexit...

Data structure

We can only observe the outcome produced by the latent variable. There are two data structures for binary data:

- classes of observations: e.g.: rats in a cage, coin tosses... $\rightarrow$ *the closest to the latent continuous variable.*
- case-based: e.g.: legislator votes, Brexit...

$\Rightarrow$ *we know the number of successes and trials in a cage/class/stratum. That's our starting point.*

Let's start with the odds

Despite binary outcomes, we want a continuous variable that is unbounded at both ends. We define a stratum and start comparing:

- Odds: Compare number of successes with number of failures within a stratum$\rightarrow$ *continuous but highly skewed.*
- Logtransform the odds $\rightarrow$ *continuous and bell shaped.*

Let's examplify with rats

We kept a 1000 rats in a cage and a number of them died (failure) while others are still alive (success). How can we model this?

We calculate the odds

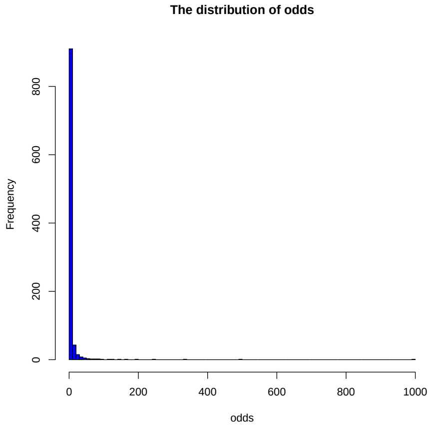
We calculate the odds of surviving in a cage in a 1000 cages

- Let's consider a series of 1000 trials where we let the successes go from complete failure (success $= 0$) to complete success (success $=$ 1000)

```r
success = 0:1000
tries = 1000
#remember: failure = tries - success
odds <- success/(tries - success)

hist(odds, breaks = 100, col = "blue")

hist(log(odds), breaks = 101, col = "blue")

plot(log(odds), success, type = "l")
```

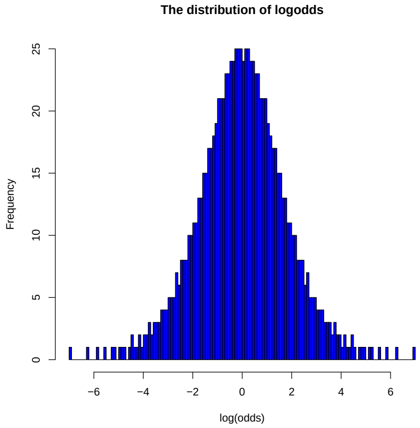
Let's start with the odds

We get a continuous but skewed variable.



Now, let's logtransform the odds

We get a nice, bellshaped curve.

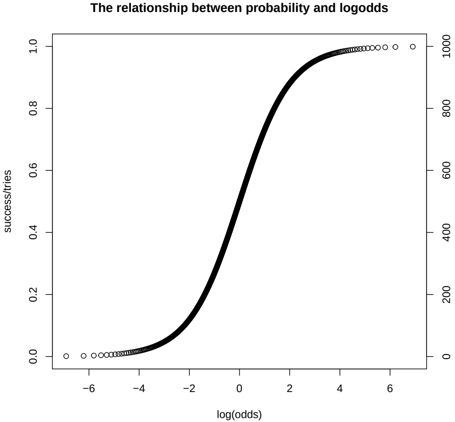


Now, let's logtransform the odds

This, we can run regressions on!

The famous S shape

We can plot the logodds of success against the number of successes or their probability (it's the same).



Probability distributions for binary variables

There are two, closely related probability distributions for binary outcomes:

- The binomial distribution: $B(n, p)$
  - p is the probability of success tells where on the x-axis (trials) the distribution is placed.
  - n is the number of trials and defines the precision (width) of the distribution.
- The Bernoulli distribution: $Ber(p)$: when we only have only one trial.

Subsection 2

Why all the fuzz? Why not OLS?

Distributions in OLS and maximum likelihood

- In OLS: The residuals must be normally distributed (but not the y_i)
- In ML: The z_i must follow a known probability distribution.

$\Rightarrow$ *This what allows us to translate the latent variable to outcomes.*

What happens if I run OLS on binary outcomes?

- The model predicts out of the possible bounderies
  - Predictions are wrong.
  - Regression coefficients are wrong.
  - Standard errors are wrong.
- The relationship between x_i and y_i is constant across all values.

$\Rightarrow$ *This last element has a bearing for the interpretation.*

Table of Contents

Section 3

Interpretation: So... what did I find?

Subsection 1

Back and forth: Logistic and logit transformation

The logit transformation

When we go from outcomes to latent variable we use the logit transformation.

$$logit(p) = log(\frac{p}{1-p}) \quad (3)$$

$\Rightarrow$ *This what R does when estimating our model*

The logistic transformation

When we go from the latent variable to outcomes we use the logistic transformation.

$$logit^{-1}(logodds) = \frac{exp(logodds)}{1 + exp(logodds)} = \frac{1}{1 + exp(-logodds)} \tag{4}$$

$\Rightarrow$ *This what we do when interpreting our model*

Subsection 2

Let's examplify with coins

Let's examplify with coins

What are the logodds for observing two heads after two tosses?

Let's examplify with coins

What are the logodds for observing two heads after two tosses?

- Probability : $Pr(y_i = 1) = \frac{1}{2} \times \frac{1}{2}$
- Odds : $Odds(y_i = 1) = \frac{Pr(y_i=1)}{1-Pr(y_i=1)}$
- Logodds: $Logodds(y_i = 1) = log(odds(y_i = 1))$

Let's examplify with coins

Here is my answer:

```
#Probability :
p = 1/4
p

## [1] 0.25

#Odds :
odds = p/(1-p)
odds

## [1] 0.3333333

#Logodds
logodds = log(odds)
logodds

## [1] -1.098612
```

Relative risk

-1.0986123, great! What does that tell us? Not much.

- But let's compare with another coin:
- We have two probabilities: Cage 1: $\frac{1}{4}$ and Cage 2: $\frac{2}{4}$
- Relative risk/odds ratio: $0.25/0.5 = 0.5$.

My three stages of interpretation

I go through tree stages of interpretation

- Inspect the marginal effects from regression table
  - Logodds: check direction and significance.
  - Odds ratio (for large coefficients) and percentage change (for smaller coefficients).
- Formulate scenarios using point estimates (in text)
- Formulate more scenarios with uncertainty using graphics.

The regression table

I interpret the regression coefficient itself

- Logodds: check direction and significance.
- Odds ratio (for large coefficients) and percentage change (for smaller coefficients).

$\Rightarrow$ A first stab to test hypotheses.

Predicted values

If you believe the model describes reality appropriately, you can learn more about it by interpreting more thoroughly

- Odds ratios are notoriously hard to understand.
- The effect depends on the value of y_i and all the other xs.

$\Rightarrow$ *Interpret the predicted values*

Formulate scenarios using point estimates (in text)

- Take an all-else-equal approach: Let one x change and keep all others constant (on a typical value).

- Find the typical representative of two x values and set the other xs accordingly.

⇒ *Which one you use depends on your objective: A theoretical point, assess effect of intervention on groups...*

Predicted values (graphic)

Formulate scenarios using point estimates and put them on speed

- Predict y values for the entire range of x and plot it.
- Simulate confidence and plot that too.
- You can do this for two scenarios.

$\Rightarrow$ *You get a sense of the actual differences in the data.*

Table of Contents

Section 4

Model assessment: How well is reality described?

Model assessment

Model assessments aim to gauge how well we describe the data (i.e. the y).

- comparison between predicted and observed values (as in OLS).
- mapping outcomes to the recoded, "latent" variable (GLM).

$\Rightarrow$ *You have a few additional "tricks" to the standard OLS assessment.*

Brier score

Describes the "average size" of the residuals.

$$B_b \equiv \frac{1}{n}\Sigma_{i=1}^{n}(\hat{\theta}_i - y_i)^2 \quad (5)$$

$\Rightarrow$ *Lower scores imply better predictions.*

How well do I discriminate?

The real question for logits is how well do I distinguish 0s from 1s.$\Rightarrow$ *Several strategies.*

Table comparison

The real question for logits is how well do I distinguish 0s from 1s.

- Table (e.g. 2×2) with proportion of predicted against observed values for 0s and 1s.
- It is χ^2 distributed (ref. the Hosmer-Lemeshow test)

$\Rightarrow$ *But how do I set the cut values (the τ)?*

The ROC curve

The ROC lets the cut values vary and displays how wrong we are on each side (true positive vs. false positive).

- A model with good predictions has a curve tending towards the upper left corner.
- The actual cut value depends on our priorities

$\Rightarrow$ *The graphic is useful in and of itself*

The separation plot

The separation plot show how the densit of observed "successes" increases as our predicted values increase.

$\Rightarrow$ *Another graphic that is useful in and of itself*