Advanced Quantitative Methods

Course for the master programme in peace and conflict studies,
Uppsala University March–May 2019

Håvard Hegre, Mihai Croicu and David Randahl

April 11, 2019

UPPSALA UNIVERSITET

The logistic regression model I

- What if the dependent variable is dichotomous?
- The logistic regression model:

$$ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1$$

  where $p = Pr(Y = 1)$
- Link function:
  the 'logit' $ln(\frac{p}{1-p})$ is the log odds of $Y = 1$
- Inverting the logit:

$$Pr(Y = 1) = p = \frac{exp(Xb)}{1 + exp(Xb)}$$

UPPSALA UNIVERSITET

The logistic regression model II

- This gives

$$Pr(Y = 1) = \frac{exp(\beta_0 + \beta_1 X_1)}{1 + exp(\beta_0 + \beta_1 X_1)}$$

- Interpretation of β_1: How much the logit (log odds of $Y = 1$) increases when X increases by one unit
- Interpretation of $exp(\beta_1)$: Odds ratio: How much the odds of $Y = 1$ increases when X increases by one unit
- It may be shown that:

  $Pr(Y_i = 0) = \frac{exp(0)}{1+exp(b_0+b_1 X_i)}$

UPPSALA UNIVERSITET

Multinomial logistic regression model I

- What if a categorical Y has more than two categories?
- Example: Conflict and resource scarcity

Scarcity	A: No conflict	B: Minor conflict	C: Major conflict
No	0.9	0.08	0.02
Yes	0.8	0.15	0.05

- Odds is the probability of something happening divided by the probability something else happens
- e.g., odds of major conflict relative to no conflict with no scarcity is $o_{CA,1} = \frac{0.02}{0.9} = 0.0222$
- odds of major conflict relative to no conflict with scarcity is $o_{CA,2} = \frac{0.05}{0.8} = 0.0625$
- The odds ratio (of major conflict relative to no conflict comparing scarcity and no scarcity): $OR_{CA} = \frac{0.0625}{0.0222} = 2.81$

UPPSALA UNIVERSITET

Multinomial logistic regression model II

- We can construct a similar set of odds and odds ratios for minor conflict versus no conflict
- We set 'no conflict' as the **reference outcome**
- If there are three values $(0, 1, 2)$ we have to say something about the probability that $Y = 1$ *as well as* $Y = 2$ relative to the probability n $pr(Y = 0)$ – *the reference outcome*
- This formulation allows us to treat the problem as two related logistic regression problems:
  - Major conflict vs. no conflict, ignoring minor conflicts
  - Minor conflict vs. no conflict, ignoring major conflicts
- The linear expression $a_0 + a_1X_i$ is related to the probability that $Y = A$
- and $b_0 + b_1X_i$ to the probability that $Y = B$

UPPSALA UNIVERSITET

Multinomial logistic regression model III

- The multinomial logistic regression model is then:

$$Pr(Y_i = A) = \frac{exp(a_0+a_1X_i)}{1+exp(a_0+a_1X_i)+exp(b_0+b_1X_i)}$$

$$Pr(Y_i = B) = \frac{exp(b_0+b_1X_i)}{1+exp(a_0+a_1X_i)+exp(b_0+b_1X_i)}$$

$$Pr(Y_i = C) = \frac{exp(0)}{1+exp(a_0+a_1X_i)+exp(b_0+b_1X_i)}$$

- This means we are estimating two models A and B, one for each comparison with the reference outcome
- Interpretation of β_As: Change in log odds of outcome A relative to outcome C when X is increased by one unit

Number of linear expressions in model

- When we estimate a multinomial model for a dependent variable with K categories, we estimate $K-1$ linear expressions (e.g. $a_0 + a_1 X_i$ and $b_0 + b_1 X_i$)
- Logistic regression is the special case where $K = 2$
- The linear expression $a_0 + a_1 X_i$ says more precisely what is the probability of $Y = A$ *relative* to that of $Y = C$
- Correspondingly, $b_0 + b_1 X_i$ models the probability of $Y = B$ relative to the probability of $Y = C$
- We do not need a corresponding expression for the probability of $Y = C$ since this is given when we know the other two
- There are *two* 'free' probabilities

UPPSALA UNIVERSITET

Conflict level as function of average income level

	Conflict level			
Income	0: None	1: Minor	2: Major	Total
Low	777 56.72 1.000	363 26.50 0.467	230 16.79 0.296	1370 100.00
Medium	2718 75.02 1.000	630 17.39 0.232	275 7.59 0.101	3623 100.00
High	2805 90.95 1.000	203 6.58 0.072	76 2.46 0.027	3084 100.00
Total	6300 78.00	1196 14.81	581 7.19	8077 100.00

Table 1: Conflict level vs income level, country years 1950–2005.
Values in cells: counts, row percentages, odds relative to no conflict

Multinomial model Conflict level vs categorical income level

Variable	Coefficient	(Std. Err.)
Equation 1 : Minor conflict		
Low income	1.865	(0.097)
Medium income	1.164	(0.085)
Intercept	-2.626	(0.073)
Equation 2 : Major conflict		
Low income	2.391	(0.138)
Medium income	1.318	(0.132)
Intercept	-3.608	(0.116)
N	8077	
Log-likelihood	-5025.239	
$\chi^2_{(4)}$	707.479	

UPPSALA UNIVERSITET

Interpretation

- Estimate for 'Low income' in 'Minor conflict' equation $a_1 = 1.865$: relative risk of being in minor vs. no conflict is $exp(1.865) = 6.5$ times higher for low-income as for high-income countries (reference category)
- I.e., $0.467/0.072 = 6.5$
- Estimate for 'Medium income' in 'Minor conflict' equation $b_1 = 1.164$: relative risk of being in minor vs. no conflict is $exp(1.164) = 3.2$ times higher for medium-income as for high-income countries
- In both equations, the estimate for 'Low income' is larger than for 'Medium income'. Possibly, the relationship is monotonically increasing with lower income levels?

De Rouen and Sobek 2004

Table I. Multinomial Logit: Civil War Terminations, 1944–97

	Outcomes			
Variable	*Government*	*Rebel*	*Truce*	*Treaty*
Bureaucracy	–.38	–7.09**	1.36	–.40
	.84	2.81	2.21	.65
Democracy	–.21	.13	–.17	.08
	.18	.27	.34	.13
Army	.26**	.31**	.57***	.29*
	.11	.14	.139	.10
Duration	–.12*	–.03	.06	–.14*
	.06	.11	.09	.06
Duration^2	.0002	–.0003	–.00008	.000
	.0002	.0006	.00028	.000
Exports	53.75*	80.59	427.00***	59.31*
	31.04	35.90	118.65	30.33
Gini	.12	.25	–2.12***	.06
	.20	.22	.66	.18
Borders	1.17**	3.20***	3.86***	.93*
	.53	1.03	1.11	.48
Ethnicity	–.03	–.18***	–.15***	–.00
	.03	.06	.05	.02
War type	–3.58	–17.92***	46.55***	–2.45
	2.51	6.99	14.37	2.43
UN	1.88	–9.76	53.44***	6.23*
	2.49	6.56	15.21	2.18
Forest	–.12*	–.57**	–.77***	–.12*
	.07	.27	.24	.07
Mountain	–.18**	.26**	.69***	–.10
	.09	.11	.24	.07
Africa	–14.86***	3.08	–37.89***	–9.92*
	4.13	5.79	9.91	3.69
Log population	1.59	.37	6.09***	.83
	.98	1.52	1.83	1.03
Log income	–5.26***	–4.78*	–14.14***	–3.78*

UPPSALA UNIVERSITET

De Rouen and Sobek reanalysis

Table 2: Simplification of Table 1, DeRouen and Sobek 2004

	Dependent variable:			
	1 Government	2 Rebel	3 Truce	4 Treaty
	(1)	(2)	(3)	(4)
Bureaucracy	0.001	−0.529	0.961	0.188
	(0.505)	(0.570)	(0.587)	(0.430)
Democracy	−0.045	−0.004	−0.037	0.075
	(0.098)	(0.105)	(0.116)	(0.092)
Army size	0.066	0.071	0.078	0.034
	(0.061)	(0.061)	(0.061)	(0.061)
Duration	−0.022***	−0.028***	−0.022**	−0.007
	(0.007)	(0.009)	(0.009)	(0.007)
Exports	2.865	−1.122	3.392	5.265
	(6.637)	(7.070)	(7.205)	(6.513)
Borders	−0.244	−0.099	−0.210	−0.167
	(0.241)	(0.254)	(0.308)	(0.227)
Log population	0.745**	0.323	0.310	−0.120
	(0.339)	(0.363)	(0.406)	(0.340)
Log income	−0.565	−0.163	−0.317	−0.514
	(0.747)	(0.778)	(0.873)	(0.724)
Constant	−4.281	−0.117	−2.956	6.396
	(5.453)	(5.961)	(6.897)	(5.778)
Akaike Inf. Crit.	277.231	277.231	277.231	277.231
Note:		*p<0.1; **p<0.05; ***p<0.01		

De Rouen and Sobek reanalysis

Table 3: DeRouen and Sobek 2004 model further simplified

	Dependent variable:			
	1 Government	2 Rebel	3 Truce	4 Treaty
	(1)	(2)	(3)	(4)
Bureaucracy	−0.041	−0.491	0.917	0.175
	(0.496)	(0.555)	(0.574)	(0.418)
Democracy	−0.014	0.011	−0.016	0.085
	(0.090)	(0.097)	(0.107)	(0.084)
Army size	0.073	0.074	0.083	0.041
	(0.061)	(0.062)	(0.062)	(0.062)
Duration	−0.023***	−0.028***	−0.022***	−0.008
	(0.007)	(0.009)	(0.008)	(0.007)
Log population	0.659**	0.358	0.214	−0.227
	(0.273)	(0.291)	(0.314)	(0.302)
Log income	−0.382	−0.106	−0.125	−0.345
	(0.659)	(0.694)	(0.773)	(0.644)
Constant	−5.184	−1.966	−3.428	6.741
	(3.258)	(3.782)	(4.777)	(4.283)
Akaike Inf. Crit.	265.526	265.526	265.526	265.526
Note:			*p<0.1; **p<0.05; ***p<0.01	

Log-likelihood: -104.7629, df=344 Table created using Stargazer (Hlavac, 2015)

UPPSALA UNIVERSITET

Significance testing

- We may test whether each parameter is different from 0 with a standard z test.
- The ratio $\frac{b_1}{se(b_1)}$ is distributed $N(0, 1)$ when N is large
- Log population is significant in the equation for government
- But is log population significant overall – does it contribute significantly to the fit of the model?

Likelihood ratio tests I

- The MLE algorithm computes a log likelihood for the estimated model
- The log likelihood is functionally similar to the χ^2 in a crosstabulation
- The exact value of the likelihood function depends on (1) data x, y, (2) the model m^* and (3) parameters θ
- If we had a larger dataset with same distribution, L would have been lower but maximum at same value
- $L(\hat{\theta}|Y, m^*) \equiv L(\hat{\theta}|Y) = k(y)Pr(Y|\hat{\theta})$
- The probability of observing these data given the model and the estimated parameters
- The logarithm of L is called *log likelihood*: LL

UPPSALA UNIVERSITET

Likelihood ratio tests II

- We can compare the log likelihood of models that are **nested**:
  - Are based on exactly the same observations
  - The nested model has fewer parameters than the outer model
  - The outer model includes all the parameters of the nested model
- The difference in log likelihood of these models multiplied by 2 has a χ^2 distribution with degrees of freedom equal to the number of parameters dropped in the nested (reduced) model
- This likelihood ratio test is useful to assess the joint impact of multiple parameters

Likelihood ratio test, DeRouen and Sobek (2004)

- Model 1: Log likelihood: -102.6144, df=336
- Model 2: Log-likelihood: -104.7629, df=344
- Difference in log likelihood: 2.1485
- $2(\ln L_1 - \ln L_2) = 4.297$
- Comparing this to the CDF of χ^2 with d.f.=8: `1-pchisq(4.297,8)` gives 0.17 (the critical value for a 95% test is 15.51 (`qchisq(0.95,8)`)
- Simplified model preferable; the Borders and Exports variables do not contribute significantly

Is the multinomal logit model ineffective?

	Conflict level			
Income	0: None	1: Minor	2: Major	Total
Low	777 56.72	363 26.50	230 16.79	1370 100.00
Medium	2718 75.02	630 17.39	275 7.59	3623 100.00
High	2805 90.95	203 6.58	76 2.46	3084 100.00
Total	6300 78.00	1196 14.81	581 7.19	8077 100.00

Table 4: Conflict level vs income level

- The effect of income stronger for major conflict than for minor?
- It makes sense to think of the conflict variable as ordered
- Can this be used to formulate a simpler model?

UPPSALA UNIVERSITET

Cumulative probabilities

- Call the conflict levels j – there are $J = 3$ categories:
  - None: $j = 1$
  - Minor: $j = 2$
  - Major: $j = 3$
- **Cumulative probabilities**:
  $P(Y \leq j) = p_1 + ... + p_j$
  - For $j = 1$: $P(Y \leq 1) = p_1$
  - For $j = 2$: $P(Y \leq 2) = p_1 + p_2$
  - For $j = 3$: $P(Y \leq 3) = p_1 + p_2 + p_3 = 1$

UPPSALA UNIVERSITET

Cumulative odds and logits

- The cumulative probabilities reflect the order of the party variable
- **Cumulative odds** for first $J-1$ categories:

$$Odds(Y>j)=\frac{P(Y>j)}{P(Y\leq j)}=\frac{1-P(Y\leq j)}{P(Y\leq j)}=\frac{p_{j+1}+...+p_J}{p_1+...+p_j}$$

- **Cumulative log odds for first $J-1$ categories**:

$$logit(Y>j)=ln(\frac{P(Y>j)}{P(Y\leq j)})=ln(\frac{p_{j+1}+...+p_J}{p_1+...+p_j})$$

Ordered models I

- Again, odds is the probability of something happening divided by the probability something else happens
- We can define this as odds of a conflict of a given intensity relative to a lesser intensity
- Two possibilities exist here (for the no scarcity row):
  - Odds of major conflict vs minor conflict or less: $o_{C,1} = \frac{0.02}{0.98} = 0.020$
  - Odds of major or minor conflict vs no conflict: $o_{C,1} = \frac{0.10}{0.9} = 0.11$
- Correspondingly for the scarcity row:
  - Odds of major conflict vs minor conflict or less: $o_{B,2} = \frac{0.05}{0.95} = 0.053$
  - Odds of major or minor conflict vs no conflict: $o_{B,2} = \frac{0.20}{0.80} = 0.25$

UPPSALA UNIVERSITET

Ordered models II

- In an ordered logistic regression model, we estimate a single odds ratio for the independent variable, assuming that it changes these two odds by the same amount
- Here, the first odds changes by $OR_C = \frac{0.053}{0.020} = 2.58 \simeq OR_B = \frac{0.25}{0.11} = 2.25$
- This smoothing allows us to estimate only one parameter per independent variable plus a set of threshold terms τ

UPPSALA UNIVERSITET

Cumulative odds and oddsratios

Scarcity	A: No conflict	B: Minor conflict	C: Major conflict
No	0.9	0.08	0.02
Yes	0.8	0.15	0.05

- Cumulative logits for no scarcity:

$$L_{01} = logit(Y > 1) = log(\frac{p_2 + p_3}{p_1}) = log(\frac{.10}{.90}) = -2.20$$

$$L_{02} = logit(Y > 2) = log(\frac{p_3}{p_1 + p_2}) = log(\frac{.02}{.98}) = -3.89$$

- Cumulative logits for scarcity:

$$L_{11} = logit(Y > 1) = log(\frac{p_2 + p_3}{p_1}) = log(\frac{.20}{.80}) = -1.39$$

$$L_{12} = logit(Y > 2) = log(\frac{p_3}{p_1 + p_2}) = log(\frac{.05}{.95}) = -2.94$$

UPPSALA UNIVERSITET

Proportional odds model

- Cumulative logits for $j = 1$:

$$0 : L_{01} = logit(Y > 1) = log(\frac{p_2 + p_3}{p_1}) = log(\frac{.10}{.90}) = -2.20$$

$$1 : L_{11} = logit(Y > 1) = log(\frac{p_2 + p_3}{p_1}) = log(\frac{.20}{.80}) = -1.39$$

- We can calculate log oddsratio for theses logits:
  $LOR_1 = L_{11} - L_{01} = -1.39 - (-2.20) = 0.81$
- Cumulative logits for $j = 2$:

$$0 : L_{02} = logit(Y > 2) = log(\frac{p_3}{p_1 + p_2}) = log(\frac{.02}{.98}) = -3.89$$

$$1 : L_{12} = logit(Y > 2) = log(\frac{p_3}{p_1 + p_2}) = log(\frac{.05}{.95}) = -2.94$$

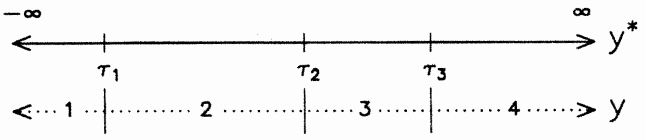
- Log oddsratio:
  $LOR_2 = L_{12} - L_{02} = -2.94 - (-3.89) = 0.95$
- $LOR_1 = 0.81 \rightarrow$ odds ratio of 2.25, $LOR_2 = 0.95 \rightarrow$ OR=2.59

Proportional odds model

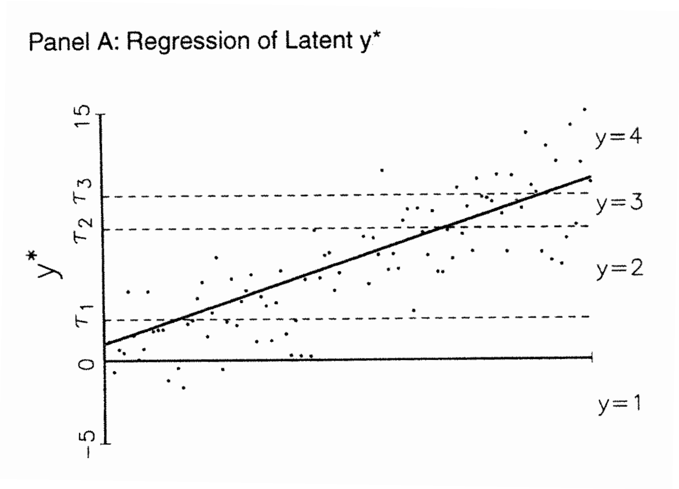
- For $j = 1$, $LOR_1 = .81$, for $j = 2$, $LOR_2 = .95$
- This is the change in cumulative logits when X increases by one unit
- It seems reasonable to assume that $LOR_1 = LOR_2 = b_1$?
- This is the same as specifying a proportional odds model: $logit(Y > j) = a_j + b_1 X; j = 1, ...J - 1$
- Note: This is a simplification relative to the contingency table

Grouped continuous dependent variable

- The dependent variable 'conflict' (none, minor, major) is a grouping of an underlying continuous variable
- We may think of conflict as a latent but unmeasured variable Y^*
- The latent variable has been captured by a categorical variable with three categories, split by two threshold values a_0 and a_1
- The latent variable might have had four categories as in the figure below, or two as in logistic regression



..


Panel A: Regression of Latent y*
y*
15
τ3
τ2
τ1
0
−5
y=4
y=3
y=2
y=1

..

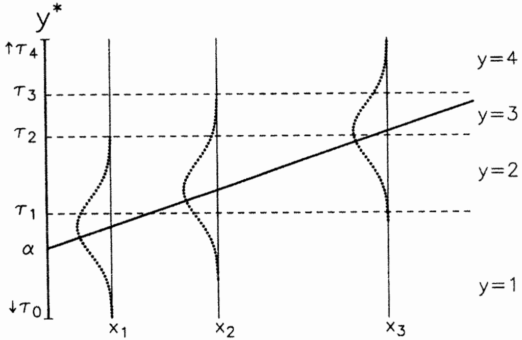
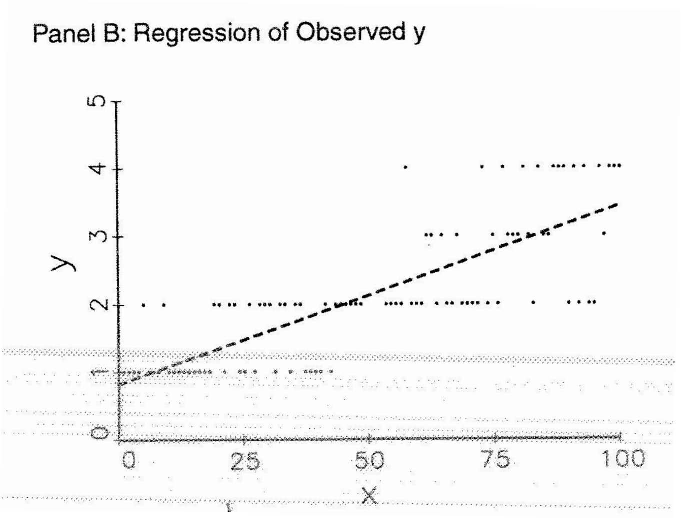


Figure 5.2. Distribution of y^* Given x for the Ordered Regression Model

..


Panel B: Regression of Observed y
y
0
2
3
4
5
0
25
50
75
100
X

UPPSALA UNIVERSITET

Ordered models I

Think of the dependent variable as a latent continuous variable, but where we observe only whether the response falls in a given category

$$Y = \begin{cases} 1 \text{ if } Y^* < \tau_1 \\ 2 \text{ if } \tau_1 \leq Y^* < \tau_2 \\ 3 \text{ if } \tau_2 \leq Y^* < \tau_3 \\ . \\ . \\ k \text{ if } Y^* \geq \tau_K \end{cases}$$

Ordered models II

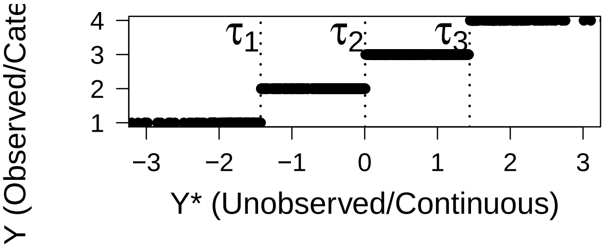


Figure 1: Illustration of the dependent variable in an ordered model

Ordered models I

- Ordered logistic regression model
  - Can be interpreted as 'cumulative odds ratios'
- Example: Probability of minor, major, or no armed conflict as a function of resource scarcity

Scarcity	A: No conflict	B: Minor conflict	C: Major conflict
No	0.9	0.08	0.02
Yes	0.8	0.15	0.05

- Ordered probit regression model
- Latent variable has a cumulative normal distribution
- The standard deviation of the assumed distribution can be adjusted to affect the scale of the τ parameters
- Interpretation analogous to logistic regression model

Example I

- Rearrange the dependent variable in DeRouen and Sobek (2004) to capture the extent to which the outcome is favorable to the rebels:
  1. Government victory
  2. Continued fighting
  3. Truce/treaty
  4. Rebel victory
- First estimate a multinomial logit version for comparison
- Then ordered logistic and probit models

Example II

Table 5: Multinomial logit, ordered recoding of Table 1, DeRouen and Sobek 2004

	Dependent variable:		
	2 Continued	3 Truce/treaty	4 Rebel victory
	(1)	(2)	(3)
Bureaucracy	0.175	0.478	−0.451
	(0.495)	(0.318)	(0.387)
Democracy	0.001	0.073	0.022
	(0.088)	(0.054)	(0.061)
Army size	−0.055	0.006	0.003
	(0.060)	(0.008)	(0.012)
Duration	0.021***	0.010**	−0.006
	(0.007)	(0.004)	(0.007)
Log population	−0.553	−0.644***	−0.309
	(0.437)	(0.217)	(0.222)
Log income	0.293	−0.034	0.268
	(0.760)	(0.407)	(0.423)
Constant	3.883	8.443*	3.431
	(9.032)	(4.880)	(5.124)
Akaike Inf. Crit.	224.243	224.243	224.243
Note:			*p<0.1; **p<0.05; ***p<0.01

Example III

Table 6: Ordered logit, ordered recoding of Table 1, DeRouen and Sobek 2004

	Dependent variable:	
	ordered.outcome	
	ordered logistic Ordered logit	*ordered probit* Ordered probit
	(1)	(2)
Bureaucracy	−0.120	−0.072
	(0.191)	(0.114)
Democracy	0.031	0.016
	(0.035)	(0.021)
Army size	0.003	0.002
	(0.005)	(0.003)
Duration	0.0003	0.00001
	(0.002)	(0.001)
Log population	−0.313**	−0.175**
	(0.135)	(0.080)
Log income	0.203	0.108
	(0.283)	(0.161)
Note:	*p<0.1; **p<0.05; ***p<0.01	

UPPSALA UNIVERSITET

Count models I

- Count data: Data that can equal (0, 1, 2, ...)
  - Number of traffic accidents in a location and/or time interval
  - Number of battle-related deaths in a country during a time period
- Poisson regresson model:

$$y_i \sim Poisson(\theta_i)$$

where

$$\theta_i = exp(X_i\beta)$$

Count models II

- Example from Gelman & Hill:

$$y_i \sim Poisson(exp(2.8 + 0.012X_{1i} - 0.20X_{2i})$$

  where X_1 is speed limit at intersection and X_2 has the value 1 if there is a traffic signal
- Constant term/intercept of model: Prediction if $X_{1i} = X_{2i} = 0$
- The coefficient of X_{1i} is the expected difference in y (on a logarithmic scale) for each additional mph of speed
- Expected multiplicative increase is $e^{0.012} = 1.012$, or a 1.2% positive difference in the rate of traffic accidents per mph

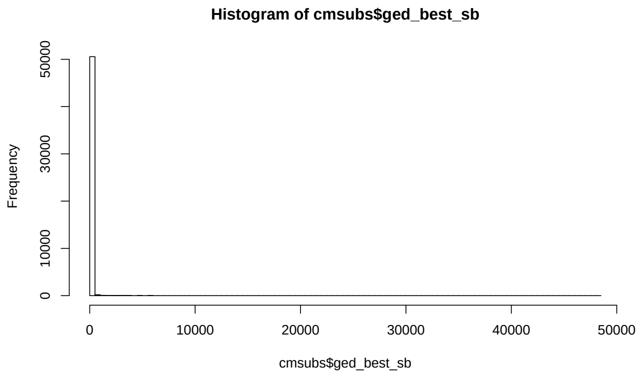
UPPSALA UNIVERSITET

Count models III

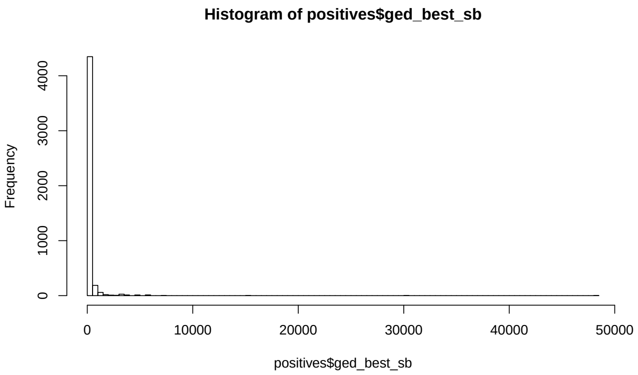
- Coefficient of X_{2i}: the predictive difference of having a traffic signal is found by multiplying the accident rate by $exp(-0.20) = 0.82$ – a reduction of 18%
- If the mean of the Poisson process is relatively high, OLS models of log counts perform well

UPPSALA UNIVERSITET

Distribution of battle-related deaths, country months, zeros included I



Distribution of battle-related deaths for months with at least 5 deaths I



UPPSALA UNIVERSITET

Overdispersion I

- Challenge: Many zeroes, high variance
- Fatalities in war may have a power-law distribution
- Under the Poisson distribution, the variance is equal to the mean
- Mathematically, $E(y_i) = u_i\theta_i$ and $sd(y_i) = \sqrt{u_i\theta_i}$
- Standardizing, residuals are

$$z_i = \frac{y_i - \hat{y}_i}{sd(\hat{y}_i)}$$

- If the Poisson model is true, the standardized residuals should have mean 0 and standard deviation 0
- If there is **overdispersion**, the standard deviation of the standardized residuals is larger than 1

Overdispersion II

- Script for testing for over-dispersion in Gelman & Hill p. 115
- Overdispersed-Poisson or Negative binomial model:

$$y_i \sim overdispersedPoisson(u_i exp(X_i \beta), \omega)$$

- where ω is the overdispersion parameter

UPPSALA UNIVERSITET

Zero-inflated models I

- Even more zeroes than in an overdispersed Poisson?
- A DGP with two separate systematic processes:
  1. A process deciding whether an observation produces a zero or a positive count
  2. A process deciding the actual count
- Variants:
  1. zero-inflated Poisson
  2. zero-inflated Negative binomial
  3. zero-inflated OLS (hurdle model)

The probit model I

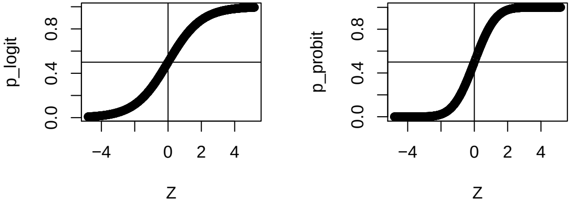


Figure 2: Link functions: logit and probit

UPPSALA UNIVERSITET

The probit model II

- The logit function maps numbers from a distribution Z ranging from $-\infty$ to ∞ to a distribution p ranging from 0 to 1
- We can think of Z as a latent variable – one that we observe only if it is larger than a given value corresponding to $p = 0.5$
- Recall that a cumulative distribution function does the same type of mapping
- The probability of observing 1 may then be formulated as coming from a normal CDF rather than the inverse logit
- Simulating the probit model:
  ```
  Y <- rbinom(n, 1, pnorm(b0 + b1*X)
  ```

The probit model III

- Estimating the probit model:
```
model <- glm(Y ~ X, family = binomial (link =
probit)
```

Bibliography I

DeRouen, Karl and David Sobek. 2004. "The Dynamics of Civil War Duration and Outcome." *Journal of Peace Research* 41(3):303–320.

Hlavac, Marek. 2015. "stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.".
URL: *http://CRAN.R-project.org/package=stargazer*