

Logistic Regression

- Logistic Regression - Dichotomous Response variable and numeric and/or categorical explanatory variable(s)
 - Goal: Model the probability of a particular as a function of the predictor variable(s)
 - Problem: Probabilities are bounded between 0 and 1
- Distribution of Responses: Binomial
- Link Function:

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

Logistic Regression with 1 Predictor

- Response - Presence/Absence of characteristic
- Predictor - Numeric variable observed for each case
- Model - $\pi(x) \equiv$ Probability of presence at predictor level x

$$\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

- $\beta = 0 \Rightarrow$ P(Presence) is the same at each level of x
- $\beta > 0 \Rightarrow$ P(Presence) increases as x increases
- $\beta < 0 \Rightarrow$ P(Presence) decreases as x increases

Logistic Regression with 1 Predictor

- α, β are unknown parameters and must be estimated using statistical software such as SPSS, SAS, or STATA
- Primary interest in estimating and testing hypotheses regarding β
 - Large-Sample test (Wald Test):
 - $H_0: \beta = 0 \quad H_A: \beta \neq 0$

$$T.S. : X_{obs}^2 = \left(\frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \right)^2$$

$$R.R. : X_{obs}^2 \geq \chi_{\alpha, 1}^2$$

$$P - val : P(\chi^2 \geq X_{obs}^2)$$

Example - Rizatriptan for Migraine

- Response - Complete Pain Relief at 2 hours (Yes/No)
- Predictor - Dose (*mg*): Placebo (0),2.5,5,10

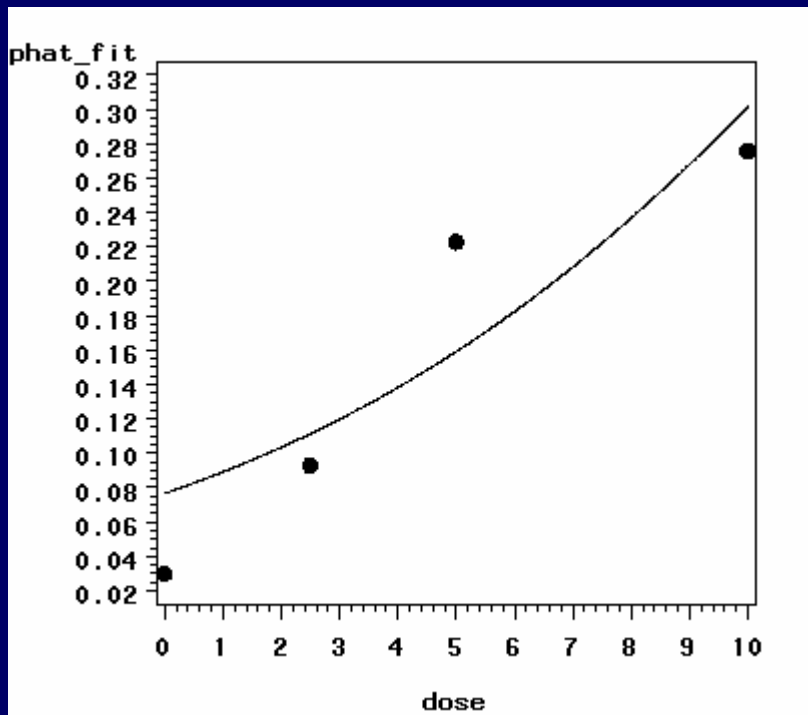
Dose	# Patients	# Relieved	% Relieved
0	67	2	3.0
2.5	75	7	9.3
5	130	29	22.3
10	145	40	27.6

Example - Rizatriptan for Migraine (SPSS)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	DOSE	.165	.037	19.819	1	.000	1.180
1	Constant	-2.490	.285	76.456	1	.000	.083

a. Variable(s) entered on step 1: DOSE.



$$\hat{\pi}(x) = \frac{e^{-2.490+0.165x}}{1 + e^{-2.490+0.165x}}$$

$$H_0 : \beta = 0 \quad H_A : \beta \neq 0$$

$$T.S.: X_{obs}^2 = \left(\frac{0.165}{0.037} \right)^2 = 19.819$$

$$RR: X_{obs}^2 \geq \chi_{.05,1}^2 = 3.84$$

$$P\text{-val} : .000$$

Odds Ratio

- Interpretation of Regression Coefficient (β):
 - In linear regression, the slope coefficient is the change in the mean response as x increases by 1 unit
 - In logistic regression, we can show that:

$$\frac{\text{odds}(x+1)}{\text{odds}(x)} = e^{\beta} \quad \left(\text{odds}(x) = \frac{\pi(x)}{1 - \pi(x)} \right)$$

- Thus e^{β} represents the change in the odds of the outcome (multiplicatively) by increasing x by 1 unit
- If $\beta = 0$, the odds and probability are the same at all x levels ($e^{\beta}=1$)
- If $\beta > 0$, the odds and probability increase as x increases ($e^{\beta}>1$)
- If $\beta < 0$, the odds and probability decrease as x increases ($e^{\beta}<1$)

95% Confidence Interval for Odds Ratio

- Step 1: Construct a 95% CI for β :

$$\hat{\beta} \pm 1.96 \hat{\sigma}_{\hat{\beta}} \equiv \left(\hat{\beta} - 1.96 \hat{\sigma}_{\hat{\beta}}, \hat{\beta} + 1.96 \hat{\sigma}_{\hat{\beta}} \right)$$

- Step 2: Raise $e = 2.718$ to the lower and upper bounds of the CI:

$$\left(e^{\hat{\beta} - 1.96 \hat{\sigma}_{\hat{\beta}}}, e^{\hat{\beta} + 1.96 \hat{\sigma}_{\hat{\beta}}} \right)$$

- If entire interval is above 1, conclude positive association
- If entire interval is below 1, conclude negative association
- If interval contains 1, cannot conclude there is an association

Example - Rizatriptan for Migraine

- 95% CI for β :

$$\hat{\beta} = 0.165 \quad \hat{\sigma}_{\hat{\beta}} = 0.037$$

$$95\% \text{ CI: } 0.165 \pm 1.96(0.037) \equiv (0.0925, 0.2375)$$

- 95% CI for population odds ratio:

$$\left(e^{0.0925}, e^{0.2375} \right) \equiv (1.10, 1.27)$$

- Conclude positive association between dose and probability of complete relief

Multiple Logistic Regression

- Extension to more than one predictor variable (either numeric or dummy variables).
- With k predictors, the model is written:

$$\pi = \frac{e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}}$$

- Adjusted Odds ratio for raising x_i by 1 unit, holding all other predictors constant:

$$OR_i = e^{\beta_i}$$

- Many models have nominal/ordinal predictors, and widely make use of dummy variables

Testing Regression Coefficients

- Testing the overall model:

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$

$$H_A : \text{Not all } \beta_i = 0$$

$$T.S. X_{obs}^2 = (-2 \log(L_0)) - (-2 \log(L_1))$$

$$R.R. X_{obs}^2 \geq \chi_{\alpha, k}^2$$

$$P = P(\chi^2 \geq X_{obs}^2)$$

- L_0, L_1 are values of the maximized likelihood function, computed by statistical software packages. This logic can also be used to compare full and reduced models based on subsets of predictors. Testing for individual terms is done as in model with a single predictor.

Example - ED in Older Dutch Men

- Response: Presence/Absence of ED ($n=1688$)
- Predictors: ($p=12$)
 - Age stratum (50-54*, 55-59, 60-64, 65-69, 70-78)
 - Smoking status (Nonsmoker*, Smoker)
 - BMI stratum (<25*, 25-30, >30)
 - Lower urinary tract symptoms (None*, Mild, Moderate, Severe)
 - Under treatment for cardiac symptoms (No*, Yes)
 - Under treatment for COPD (No*, Yes)

* Baseline group for dummy variables

Example - ED in Older Dutch Men

Predictor	b	s _b	Adjusted OR (95% CI)
Age 55-59 (vs 50-54)	0.83	0.42	2.3 (1.0 – 5.2)
Age 60-64 (vs 50-54)	1.53	0.40	4.6 (2.1 – 10.1)
Age 65-69 (vs 50-54)	2.19	0.40	8.9 (4.1 – 19.5)
Age 70-78 (vs 50-54)	2.66	0.41	14.3 (6.4 – 32.1)
Smoker (vs nonsmoker)	0.47	0.19	1.6 (1.1 – 2.3)
BMI 25-30 (vs <25)	0.41	0.21	1.5 (1.0 – 2.3)
BMI >30 (vs <25)	1.10	0.29	3.0 (1.7 – 5.4)
LUTS Mild (vs None)	0.59	0.41	1.8 (0.8 – 4.3)
LUTS Moderate (vs None)	1.22	0.45	3.4 (1.4 – 8.4)
LUTS Severe (vs None)	2.01	0.56	7.5 (2.5 – 22.5)
Cardiac symptoms (Yes vs No)	0.92	0.26	2.5 (1.5 – 4.3)
COPD (Yes vs No)	0.64	0.28	1.9 (1.1 – 3.6)

Interpretations: Risk of ED appears to be:

- Increasing with age, BMI, and LUTS strata
- Higher among smokers
- Higher among men being treated for cardiac or COPD

Loglinear Models with Categorical Variables

- Logistic regression models when there is a clear response variable (Y), and a set of predictor variables (X_1, \dots, X_k)
- In some situations, the variables are all responses, and there are no clear dependent and independent variables
- Loglinear models are to correlation analysis as logistic regression is to ordinary linear regression

Loglinear Models

- Example: 3 variables (X, Y, Z) each with 2 levels
- Can be set up in a $2 \times 2 \times 2$ contingency table
- Hierarchy of Models:
 - All variables are conditionally independent
 - Two of the pairs of variables are conditionally independent
 - One of the pairs are conditionally independent
 - No pairs are conditionally independent, but each association is constant across levels of third variable (**no interaction** or **homogeneous association**)
 - All pairs are associated, and associations differ among levels of third variable

Loglinear Models

- To determine associations, must have a measure: the odds ratio (OR)
- Odds Ratios take on the value 1 if there is no association
- Loglinear models make use of regressions with coefficients being exponents. Thus, tests of whether odds ratios are 1, is equivalently to testing whether regression coefficients are 0 (as in logistic regression)
- For a given partial table, $OR=e^{\beta}$, software packages estimate and test whether $\beta=0$

Example - Feminine Traits/Behavior

- 3 Variables, each at 2 levels (Table contains observed counts):
- Feminine **P**ersonality Trait (Modern/Traditional)
 - Female **R**ole Behavior (Modern/Traditional)
 - C**lass (Lower Classman/Upper Classman)

PRSNALTY * ROLEBHVR * CLASS1 Crosstabulation

Count

CLASS1			ROLEBHVR		Total
			Modern	Traditional	
Lower Classman	PRSNALTY	Modern	33	25	58
		Traditional	21	53	74
	Total		54	78	132
Upper Classman	PRSNALTY	Modern	19	13	32
		Traditional	10	35	45
	Total		29	48	77

Example - Feminine Traits/Behavior

- Expected cell counts under model that allows for association among all pairs of variables, but no interaction (association between personality and role is same for each class, etc).

Model:(*PR,PC,RC*)

- Evidence of personality/role association (see odds ratios)

	Class=Lower	Class=Lower	Class=Upper	Class=Upper
	Role=M	Role=T	Role=M	Role=T
Personality=M	34.1	23.9	17.9	14.1
Personality=T	19.9	54.1	11.1	33.9

$$\text{Class = Lower : } OR = \frac{34.1(54.1)}{23.9(19.9)} = 3.88$$

$$\text{Class = Upper : } OR = \frac{17.9(33.9)}{14.1(11.1)} = 3.88$$

Note that under the no interaction model, the odds ratios measuring the personality/role association is same for each class

Example - Feminine Traits/Behavior

	Role=M	Role=M	Role=T	Role=T
	Class=Lower	Class=Upper	Class=Lower	Class=Upper
Personality=M	34.1	17.9	23.9	14.1
Personality=T	19.9	11.1	54.1	33.9

$$\text{Role} = \text{M} : OR = \frac{34.1(11.1)}{17.9(19.9)} = 1.06$$

$$\text{Role} = \text{T} : OR = \frac{23.9(33.9)}{14.1(54.1)} = 1.06$$

	Personality=M	Personality=M	Personality=T	Personality=T
	Class=Lower	Class=Upper	Class=Lower	Class=Upper
Role=M	34.1	17.9	19.9	11.1
Role=T	23.9	14.1	54.1	33.9

$$\text{Personality} = \text{M} : OR = \frac{34.1(14.1)}{17.9(23.9)} = 1.12$$

$$\text{Personality} = \text{T} : OR = \frac{19.9(33.9)}{11.1(54.1)} = 1.12$$

Example - Feminine Traits/Behavior

- Intuitive Results:
 - Controlling for class in school, there is an association between personality trait and role behavior ($OR_{Lower} = OR_{Upper} = 3.88$)
 - Controlling for role behavior there is no association between personality trait and class ($OR_{Modern} = OR_{Traditional} = 1.06$)
 - Controlling for personality trait, there is no association between role behavior and class ($OR_{Modern} = OR_{Traditional} = 1.12$)

SPSS Output

- Statistical software packages fit regression type models, where the regression coefficients for each model term are the log of the odds ratio for that term, so that the estimated odds ratio is e raised to the power of the regression coefficient.

Parameter Estimates

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
Constant	3.5234	.1651	21.35	3.20	3.85
Class	.4674	.2050	2.28	.07	.87
Personality	-.8774	.2726	-3.22	-1.41	-.34
Role	-1.1166	.2873	-3.89	-1.68	-.55
C*P	.0605	.3064	.20	-.54	.66
C*R	.1166	.3107	.38	-.49	.73
R*P	1.3554	.2987	4.54	.77	1.94

Note: $e^{1.3554} = 3.88$ $e^{.0605} = 1.06$ $e^{.1166} = 1.12$

Interpreting Coefficients

- The regression coefficients for each variable corresponds to the lowest level (in alphanumeric ordering of symbols). Computer output will print a “mapping” of coefficients to variable levels.

Cell (C,P,R)	Class β	Prsnlty β	Role β	C*P β	C*R β	P*R β	Expected Count
L,M,M	0.4674	-0.8774	-1.1166	0.0605	0.1166	1.3554	34.1
L,M,T	0.4674	-0.8774	0	0.0605	0	0	23.9
L,T,M	0.4674	0	-1.1166	0	0.1166	0	19.9
L,T,T	0.4674	0	0	0	0	0	54.1
U,M,M	0	-0.8774	-1.1166	0	0	1.3554	17.9
U,M,T	0	-0.8774	0	0	0	0	14.1
U,T,M	0	0	-1.1166	0	0	0	11.1
U,T,T	0	0	0	0	0	0	33.9

To obtain the expected cell counts, add the constant (3.5234) to each of the β s for that row, and raise e to the power of that sum

Goodness of Fit Statistics

- For any logit or loglinear model, we will have contingency tables of observed (f_o) and expected (f_e) cell counts under the model being fit.
- Two statistics are used to test whether a model is appropriate: the Pearson chi-square statistic and the likelihood ratio (aka Deviance) statistic

$$\text{Pearson Chi - square : } \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\text{Likelihood - Ratio : } G^2 = 2 \sum f_o \log \left(\frac{f_o}{f_e} \right)$$

Goodness of Fit Tests

- Null hypothesis: The current model is appropriate
- Alternative hypothesis: Model is more complex
- Degrees of Freedom: Number of cells-Number of parameters in model
- Distribution of Goodness of Fit statistics under the null hypothesis is chi-square with degrees of freedom given above
- Statistical software packages will print these statistics and P-values.

Example - Feminine Traits/Behavior

Table Information

Factor	Value	Observed Count	%	Expected Count	%
PRSNALTY	Modern				
ROLEBHVR	Modern				
CLASS1	Lower Classman	33.00	(15.79)	34.10	(16.32)
CLASS1	Upper Classman	19.00	(9.09)	17.90	(8.57)
ROLEBHVR	Traditional				
CLASS1	Lower Classman	25.00	(11.96)	23.90	(11.44)
CLASS1	Upper Classman	13.00	(6.22)	14.10	(6.75)
PRSNALTY	Traditional				
ROLEBHVR	Modern				
CLASS1	Lower Classman	21.00	(10.05)	19.90	(9.52)
CLASS1	Upper Classman	10.00	(4.78)	11.10	(5.31)
ROLEBHVR	Traditional				
CLASS1	Lower Classman	53.00	(25.36)	54.10	(25.88)
CLASS1	Upper Classman	35.00	(16.75)	33.90	(16.22)

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	.4695	1	.4932
Pearson	.4664	1	.4946

Example - Feminine Traits/Behavior

Goodness of fit statistics/tests for all possible models:

Model	G^2	χ^2	df	P-value (G^2)	P-value (χ^2)
(C,P,R)	22.21	22.46	4	.0002	.0002
(C,PR)	0.7199	0.7232	3	.8685	.8677
(P,CR)	21.99	22.24	3	.00007	.00006
(R,CP)	22.04	22.34	3	.00006	.00006
(CR,CP)	22.93	22.13	2	.00002	.00002
(CP,PR)	0.6024	0.6106	2	.7399	.7369
(CR,PR)	0.5047	0.5085	2	.7770	.7755
(CP,CR,PR)	0.4644	0.4695	1	.4946	.4932

The simplest model for which we fail to reject the null hypothesis that the model is adequate is: (C,PR): Personality and Role are the only associated pair.

Adjusted Residuals

- Standardized differences between actual and expected counts ($f_o - f_e$, divided by its standard error).
- Large adjusted residuals (bigger than 3 in absolute value, is a conservative rule of thumb) are cells that show lack of fit of current model
- Software packages will print these for logit and loglinear models

Example - Feminine Traits/Behavior

- Adjusted residuals for (C,P,R) model of all pairs being conditionally independent:

Factor	Value	Resid.	Adj. Resid.
PRSNALTY	Modern		
ROLEBHVR	Modern		
CLASS1	Lower Classman	10.43	3.04**
CLASS1	Upper Classman	5.83	1.99
ROLEBHVR	Traditional		
CLASS1	Lower Classman	-9.27	-2.46
CLASS1	Upper Classman	-6.99	-2.11
PRSNALTY	Traditional		
ROLEBHVR	Modern		
CLASS1	Lower Classman	-8.85	-2.42
CLASS1	Upper Classman	-7.41	-2.32
ROLEBHVR	Traditional		
CLASS1	Lower Classman	7.69	1.93
CLASS1	Upper Classman	8.57	2.41

Comparing Models with G^2 Statistic

- Comparing a series of models that increase in complexity.
- Take the difference in the deviance (G^2) for the models (less complex model minus more complex model)
- Take the difference in degrees of freedom for the models
- Under hypothesis that less complex (reduced) model is adequate, difference follows chi-square distribution

Example - Feminine Traits/Behavior

- Comparing a model where only Personality and Role are associated (Reduced Model) with the model where all pairs are associated with no interaction (Full Model).
- Reduced Model (C,PR): $G^2=.7232$, $df=3$
- Full Model (CP,CR,PR): $G^2=.4695$, $df=1$
- Difference: $.7232-.4695=.2537$, $df=3-1=2$
- Critical value ($\alpha=0.05$): 5.99
- Conclude Reduced Model is adequate

Logit Models for Ordinal Responses

- Response variable is ordinal (categorical with natural ordering)
- Predictor variable(s) can be numeric or qualitative (dummy variables)
- Labeling the ordinal categories from 1 (lowest level) to c (highest), can obtain the **cumulative probabilities**:

$$P(Y \leq j) = P(Y = 1) + \cdots + P(Y = j) \quad j = 1, \dots, c$$

Logistic Regression for Ordinal Response

- The odds of falling in category j or below:

$$\frac{P(Y \leq j)}{P(Y > j)} \quad j = 1, \dots, c-1 \quad P(Y \leq c) = 1$$

- Logit (log odds) of cumulative probabilities are modeled as linear functions of predictor variable(s):

$$\text{logit}[P(Y \leq j)] = \log \left[\frac{P(Y \leq j)}{P(Y > j)} \right] = \alpha_j + \beta X \quad j = 1, \dots, c-1$$

This is called the **proportional odds model**, and assumes the effect of X is the same for each cumulative probability

Example - Urban Renewal Attitudes

- Response: Attitude toward urban renewal project (Negative ($Y=1$), Moderate ($Y=2$), Positive ($Y=3$))
- Predictor Variable: Respondent's Race (White, Nonwhite)
- Contingency Table:

Attitude\Race	White	Non white
Negative ($Y=1$)	101	106
Moderate ($Y=2$)	91	127
Positive ($Y=3$)	170	190

SPSS Output

- Note that SPSS fits the model in the following form:

$$\text{logit}[P(Y \leq j)] = \log \left[\frac{P(Y \leq j)}{P(Y > j)} \right] = \alpha_j - \beta X \quad j = 1, \dots, c-1$$

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[ATTITUDE = 1]	-1.027	.102	101.993	1	.000	-1.227	-.828
	[ATTITUDE = 2]	.165	.094	3.070	1	.080	-.020	.351
Location	[RACE=0]	-.001	.133	.000	1	.993	-.263	.260
	[RACE=1]	0 ^a	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Note that the race variable is not significant (or even close).

Fitted Equation

- The fitted equation for each group/category:

$$\text{Negative/White : logit} \left[\frac{P(Y \leq 1 | \text{White})}{P(Y > 1 | \text{White})} \right] = -1.027 - (-0.001) = -1.026$$

$$\text{Negative/Nonwhite : logit} \left[\frac{P(Y \leq 1 | \text{Nonwhite})}{P(Y > 1 | \text{Nonwhite})} \right] = -1.027 - (0) = -1.027$$

$$\text{Neg or Mod/White : logit} \left[\frac{P(Y \leq 2 | \text{White})}{P(Y > 2 | \text{White})} \right] = 0.165 - (-0.001) = 0.166$$

$$\text{Neg or Mod/Nonwhite : logit} \left[\frac{P(Y \leq 2 | \text{Nonwhite})}{P(Y > 2 | \text{Nonwhite})} \right] = 0.165 - 0 = 0.165$$

For each group, the fitted probability of falling in that set of categories is $e^L/(1+e^L)$ where L is the logit value (0.264,0.264,0.541,0.541)

Inference for Regression Coefficients

- If $\beta = 0$, the response (Y) is independent of X
- Z-test can be conducted to test this (estimate divided by its standard error)
- Most software will conduct the Wald test, with the statistic being the z-statistic squared, which has a chi-squared distribution with 1 degree of freedom under the null hypothesis
- Odds ratio of increasing X by 1 unit and its confidence interval are obtained by raising e to the power of the regression coefficient and its upper and lower bounds

Example - Urban Renewal Attitudes

Parameter Estimates

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Threshold [ATTITUDE = 1]	-1.027	.102	101.993	1	.000	-1.227	-.828
[ATTITUDE = 2]	.165	.094	3.070	1	.080	-.020	.351
Location [RACE=0]	-.001	.133	.000	1	.993	-.263	.260
[RACE=1]	0 ^a	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

- Z-statistic for testing for race differences:
 $Z = 0.001 / 0.133 = 0.0075$ (recall model estimates $-\beta$)
- Wald statistic: .000 (P -value=.993)
- Estimated odds ratio: $e^{.001} = 1.001$
- 95% Confidence Interval: $(e^{-.260}, e^{.263}) = (0.771, 1.301)$
- Interval contains 1, odds of being in a given category or below is same for whites as nonwhites

Ordinal Predictors

- Creating dummy variables for ordinal categories treats them as if nominal
- To make an ordinal variable, create a new variable X that models the levels of the ordinal variable
- Setting depends on assignment of levels (simplest form is to let $X=1, \dots, c$ for the categories which treats levels as being equally spaced)