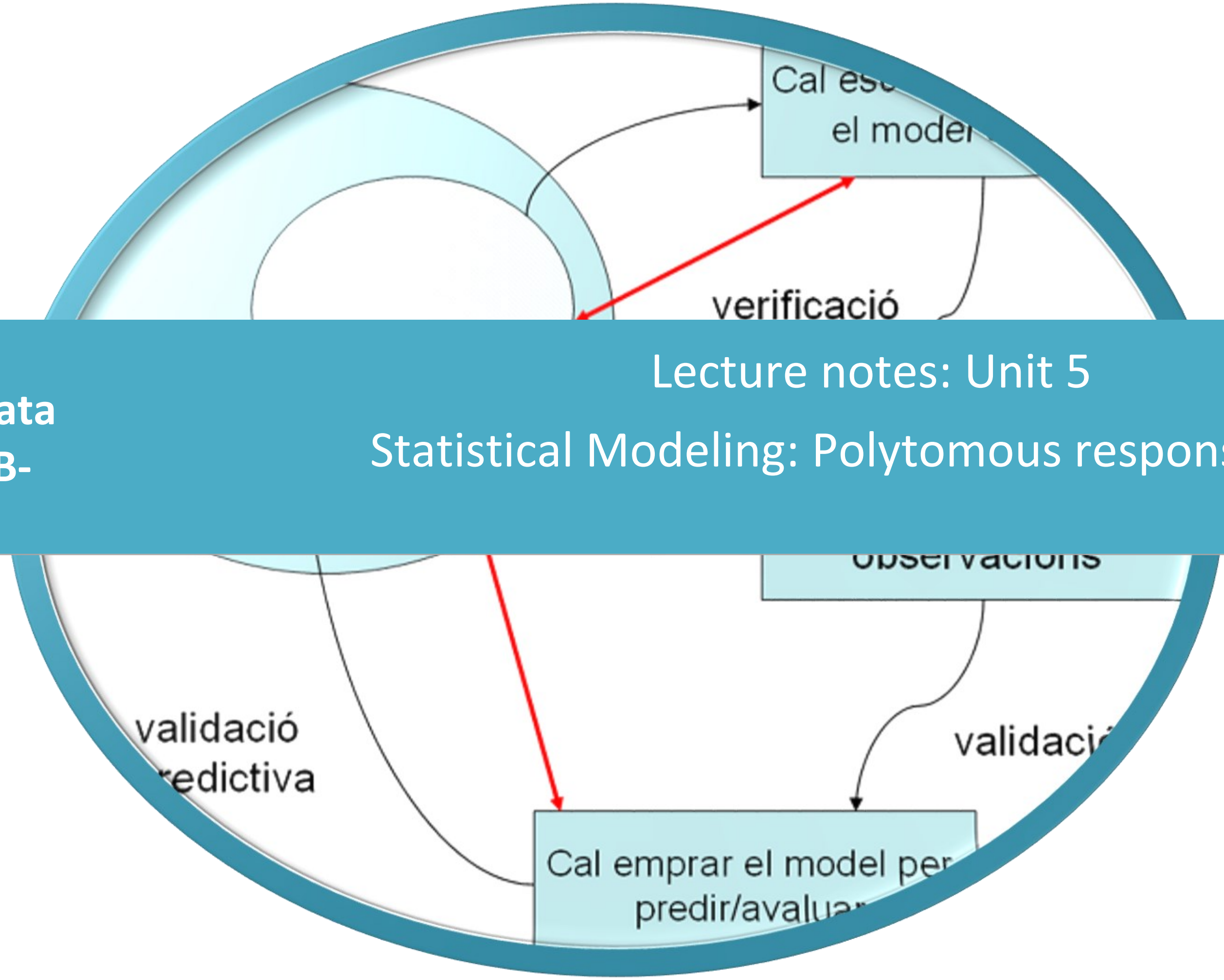
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

2024-2025 Academic Year



SIM course.
Master in Data Science – FIB-UPC

Lecture notes: Unit 5

Statistical Modeling: Polytomous response data

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística i Investigació Operativa

TABLE OF CONTENTS

5-1. POLYTOMOUS RESPONSE DATA. MULTINOMIAL MODELS

5-1.1 Components of generalized linear models

Generalized linear models are extensions of classic multiple regression models.

Let $\mathbf{y}^T = (\mathbf{y_1}, \ldots, \mathbf{y_n})$ be a vector of n components randomly drawn from vector $\mathbf{Y}^T = (\mathbf{Y_1}, \ldots, \mathbf{Y_n})$, whose variables are statistically independent and distributed with expectation $\mu^T = (\mu_1, \ldots, \mu_n)$:

The random component assumes that mutual independence holds and each random variable in $\mathbf{Y}^T = (\mathbf{Y_1}, \ldots, \mathbf{Y_n})$ belongs to the exponential family with one parameter distribution $\boldsymbol{Y_i}|X_i \sim F(.\,; \boldsymbol{\mu_i}, \phi)$ and expected values $E(\boldsymbol{Y}_i|X_i) = \boldsymbol{\mu_i}$.

➡ **For grouped data for each observation group, the response is polytomous and we are dealing with multinomial distribution.**

- **The systematic component in the model** specifies a vector $\boldsymbol{\eta}$. The linear predictor vector is a linear combination from a limited number of explanatory variables $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_p})$ or regressors and parameters $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_p)$ to be estimated. In matrix notation, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ where $\boldsymbol{\eta}$ is *nx1*, **X** is *nxp* and $\boldsymbol{\beta}$ is *px1*.

POLYTOMOUS RESPONSE DATA. MULTINOMIAL MODELS

For each observation i, the expected value $\boxed{\boldsymbol{\pi}_{\mathbf{i}}^{\mathbf{T}} = (\pi_{i1} \quad \cdots \quad \pi_{iK})}$ is related to the linear predictor $\boxed{\boldsymbol{\eta}_{\mathbf{i}}}$ through the scalar ***link function*, denoted *g(.)***, and thus $g(\mu_{ik}) = \mathbf{X}_{\mathbf{ik}}^{\mathbf{T}}\boldsymbol{\beta} = \eta_{ik} \quad \forall k \in 1 \cdots K$.

The response function is $\boldsymbol{\mu}_{\mathbf{i}} = \mathbf{g}^{-1}\left(\mathbf{X}_{\mathbf{i}}^{\mathbf{T}}\boldsymbol{\beta}\right) = \mathbf{g}^{-1}(\boldsymbol{\eta}_{\mathbf{i}})$

In ordinary least squares models for normal data, the identity link used is $\boldsymbol{\eta} = \boldsymbol{\mu}$.

For polytomous data, several treatments are commonly used and will be presented in a later section. Basically, a log transformation once one category in the outcome is chosen as a reference.

Since ML estimates: $\boxed{\widehat{\boldsymbol{\beta}} \ \forall i \forall k \rightarrow \hat{\eta}_{ik} = \mathbf{X}_{\mathbf{ik}}^{\mathbf{T}}\widehat{\boldsymbol{\beta}} \ \rightarrow \hat{\pi}_{ik} = g^{-1}(\hat{\eta}_{ik}) \rightarrow \hat{\mu}_{ik} = m_i\hat{\pi}_{ik}}$

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

5-2. INTRODUCTION TO POLYTOMOUS TARGET MODELLING

On Unit 4, binary models were presented. These models are rather limited since a dichotomous variable is used to model the target distribution. In practical studies, polytomous data arise and polytomous target variables have to be considered, either from a nominal, or ordinal point of view.

➡ polytomous target data modelling has been under from 3 different approaches:

1. Directly modelling polytomous targets as a nominal categorical variable without any order between levels, being a generalization of binary response models. Ordinal categories are lost.
2. By extending binary target model to hierarchical binary decisions, where each dichotomic decision is considered separately.
3. Modelling ordinal polytomous target explicitly based on a latent variable approach. An unobserved propension concept is involved in this framework.

➡ McCullagh book offers a unified approach to polytomous data modelling, very consistent, but tough.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

INTRODUCTION TO POLYTOMOUS TARGET MODELLING

5-2.1 Multinomial distribution

This is the more natural distribution to address polytomous target variables.

➡ A multinomial distribution arises as a consequence of sampling in a population where each unit belong to 1 and only 1 category from 1 to K, $A_1 \ldots A_K$. Exhaustive coverage of the population is needed.

If sample size is m and population size is infinite (or very large, over 500000 units), then a random sampling of size $\boxed{m}$ (units) drawing $\boxed{y_k}$ units belonging to category *k*, for k in 1...K, determined by a probability $\boxed{\pi_k}$ according to the multinomial probability function:

$$P\left(Y_1 = y_1, \ldots, Y_K = y_K, \pi\right) = \binom{m}{\mathbf{y}} \pi_1^{y_1} \ldots \pi_K^{y_K}$$

where π is interpreted as the vector of probabilities and $\boxed{y_k}$ as the vector of sample frequencies and

$$\binom{m}{\mathbf{y}} = \frac{m!}{y_1! \ldots y_K!} \text{ y } 0 \leq y_j \leq m \text{ s.t. } \sum_j y_j = m .$$

INTRODUCTION TO POLYTOMOUS TARGET MODELLING

➡ An alternative derivation of the multinomial vector assumes $Y_1,\ldots,Y_k$ K independent random variables Poisson distributed with expected means $\mu_1,\ldots,\mu_k$, and thus, the conditional joint distribution of $Y_1,\ldots,Y_k$ given

$$Y_+ = \sum_k Y_k \text{ is } \mathrm{P}\left(Y_1 = y_1,\ldots,Y_k = y_k,\boldsymbol{\pi}\right) = \binom{m}{\mathbf{y}} \pi_1^{y_1} \ldots \pi_k^{y_k}$$

$$\text{where } \pi_j = \frac{\mu_j}{\mu_+} \text{ and } \mu_+ = \sum_j \mu_j .$$

➡ Basic moments (expectation and variance) are:

- $$\mathrm{E}[\mathbf{Y}] = \begin{pmatrix} m\pi_1 \\ \vdots \\ m\pi_k \end{pmatrix} \text{ and } \Sigma = \mathbf{V}[\mathbf{Y}] = m\left\{diag(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T\right\} = \begin{cases} m\pi_r(1-\pi_r) & r = s \\ -m\pi_r\pi_s & r \neq s \end{cases}.$$

Variance-covariance matrix Σ $\boxed{KxK}$ has $\boxed{K\text{-}1}$ range, leading to some numerical issues when applying a ML estimation (Moore-Penrose pseudo-inverse is the best option).

INTRODUCTION TO POLYTOMOUS TARGET MODELLING

Some interesting properties are:

- Marginal distribution for each category is binomial distributed: $\boxed{Y_j \approx B(m, \pi_j)}$.
- Joint marginal distribution of 2 components of the multinomial vector, let us say $\boxed{\textbf{r and s}}$ follows a 3-way multinomial distribution being $\boxed{\textbf{m}}$ index and probability parameters $(\pi_r \quad \pi_s \quad 1-\pi_r-\pi_s)$. This property can be extended to more than 2 components.
- Joint conditional distribution of $Y_1, \ldots, Y_K$ *given* $\boxed{Y_j = y_j}$ is **multinomial** distributed with K-1 categories (*j is removed*), index vector $m - y_j$ and parameters (normalized probabilities),

$$\boxed{\pi_r \leftarrow \frac{\pi_r}{1-\pi_j}} \qquad 1 \le r \le K \quad r \ne j$$

- Hypothesis testing H$_0$: $\boxed{\boldsymbol{\pi} = \boldsymbol{\pi}_o}$ using Pearson statistic involves a quadratic form definition:

$$\boxed{\mathbf{X}_\mathbf{P}^2 = \mathbf{R}^\mathbf{T}\Sigma^-\mathbf{R} \qquad R_j = y_j - m\pi_{0j} \qquad \sum R_j = 0}$$

(any pseudo-inverse might be used).

If $\boxed{\textbf{m}}$ is large then $\mathbf{X}_\mathbf{P}^2 \approx \chi_{K-1}^2$.

INTRODUCTION TO POLYTOMOUS TARGET MODELLING

- Ordinal polytomous target approaches rely on defining the cumulative multinomial vector and the cumulative multinomial distribution, ***vector of cumulative probabilities of*** $\boxed{\pi}$, usually notated as $\boxed{\gamma}$, and defined as, $\gamma_j = \sum_{r \le j} \pi_r \quad 1 \le j \le K-1 \quad \gamma_K = 1$.

$$\mathbf{Z} = \mathbf{LY} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ \vdots \\ \vdots \\ Y_K \end{bmatrix}$$

- Introducing some algebraic perspective, the new vector variable **Z**, cumulative multinomial vector, is a lineal function of **Y** original multinomial vector: **Z=LY**.
- Basic moment for **Z** vector are:

$$\mathrm{E}[\mathbf{Z}] = \begin{pmatrix} m\gamma_1 \\ \vdots \\ m\gamma_K \end{pmatrix}$$ and

$$\Sigma_Z = \mathbf{V}[\mathbf{Z}] = \begin{cases} m\gamma_r(1-\gamma_s) & r \le s \\ 0 & \textit{sinó} \end{cases}$$ upper triangular of range $\boxed{\textbf{\textit{K-1}}}$.

The cumulative multinomial vector **Z** has nicer properties than the original **Y** vector.

5-3. POLYTOMOUS TARGET MODELLING: LOG-LIKELIHOOD

5-3.1 Loglikelihood function for multinomial data

Observed data are vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ **n independent multinomial vectors**, each of them with **K categories** where $\mathbf{y}_i^T = \begin{pmatrix} y_{i1} & \ldots & y_{iK} \end{pmatrix}$ and $\boxed{\sum_j y_{ij} = m_i}$.

➡ Deviance and scaled deviance are identical:

$$D(\mathbf{y}, \hat{\mu}) = D'(\mathbf{y}, \hat{\mu})\phi = D'(\mathbf{y}, \hat{\mu}) = 2\,\ell(\mathbf{y}, \mathbf{y}) - 2\,\ell(\hat{\mu}, \mathbf{y}).$$

➡ Saturated model $\ell(\mathbf{y}, \mathbf{y})$ (as many parameters as observations by categories) assumes $\boxed{\tilde{\pi}_{ij} = \frac{y_{ij}}{m_i}}$, and it is notated $\ell(\tilde{\pi}, \mathbf{y})$.

$$\boxed{\begin{aligned} &D'(\mathrm{y}, \hat{\pi}) = 2\,\ell(\tilde{\pi}, \mathrm{y}) - 2\,\ell(\hat{\pi}, \mathrm{y}) = \\ &= 2\sum_{i,j} y_{ij} \log \tilde{\pi}_{ij} - 2\sum_{i,j} y_{ij} \log \hat{\pi}_{ij} = 2\sum_{i,j} y_{ij} \log \frac{y_{ij}}{m_i \hat{\pi}_{ij}} = 2\sum_{i,j} y_{ij} \log \frac{y_{ij}}{\hat{\mu}_{ij}} \end{aligned}}$$

➡ Under some large size index conditions and $\hat{\mu}_{ij}$ large without overdispersion, then deviance is assimptotically distributed as a chi squared law χ^2.

5-4. POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

➡ As indicated before, several approaches are considered:

I. **Nominal scales**, where the categories are nominal categories.

II. **Ordinal scales,** where ordered categories are present and they correspond to an ordinal scale first, second, etc. It is makes no sense to consider a distance among pairs of categories, just ordering is meaningful.

III. **Interval scales,** where ordered categories are present and they correspond to numeric labels usually being central points in the interval. Differences among categories are meaningful. Nevertheless, suitable models will not be addressed in the current course.

- Distinction among ordinal and interval scales is not always evident. For example: a market study leading to establish trade mark preferences can be modelled as an ordinal scale when a Likert-like scale is used to describe categories: excellent, good, medium, bad, horrible.
- Nevertheless, responses related to media preferences or political parties can be modelled as either an ordinal type, or as an interval scale type.

POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

5-4.1 Models for nominal scales

Probability vector π is directly used, γ not considered. **Logarithmic link allows to decompose the linear predictor into as many effects as explanatory variables are considered.**

We have to get used to the base-line reparametrization to simplify model interpretation $\eta_{i1}(\mathbf{x_i}) = 0$ base-line category 1. Models expressed in log-odds terms with respect to base-line category 1 behave as:

- No covariate effects, null model (1):

$$\eta_{ij}(\mathbf{x_i}) = \log \frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})} = \alpha_j \quad j = 2,\ldots,K \quad i = 1,\ldots,n$$

- Model (X), conditional logit models can be estimated using mlogit package in R:

$$\eta_{ij}(\mathbf{x_i}) = \log \frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})} = \alpha_j + \beta^{\mathbf{T}} \mathbf{x_i} \quad j = 2,\ldots,K \quad i = 1,\ldots,n$$

- Model (X), the ones estimated using R by default (for MINITAB/R reference is K):

$$\eta_{ij}(\mathbf{x_i}) = \log \frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})} = \alpha_j + \beta_j^{\mathbf{T}} \mathbf{x_i} \quad j = 2,\ldots,K \quad i = 1,\ldots,n$$

POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

- Odds of jth category over baseline category (1st category) becomes,

$$\frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})} = \exp\left(\alpha_j + \beta_j^{\mathrm{T}} \mathbf{x_i}\right) = \exp\left(\eta_{ij}(\mathbf{x}_i)\right) \quad j = 2, \ldots, K \qquad i = 1, \ldots, n$$

- Odds of jth category over l-th category becomes, $j \neq 1, l \neq 1$,

$$\boxed{\frac{\pi_{ij}(\mathbf{x_i})}{\pi_{il}(\mathbf{x_i})} = \frac{\pi_{ij}(\mathbf{x_i})/\pi_{i1}(\mathbf{x_i})}{\pi_{il}(\mathbf{x_i})/\pi_{i1}(\mathbf{x_i})} = \exp\left\{\left(\alpha_j - \alpha_l\right) + \left(\beta_j - \beta_l\right)^{\mathrm{T}} \mathbf{x_i}\right\}}$$

According to base-line reparametrization, $\begin{matrix} j = 1, \cdots, K \\ i = 1, \cdots, n \end{matrix}$

$$\boxed{\pi_{ij}(\mathbf{x}_i) = \frac{\mathbf{exp}\left(\eta_{ij}(\mathbf{x}_i)\right)}{\sum_r \mathbf{exp}\left(\eta_{ir}(\mathbf{x}_i)\right)}}$$

and

$$\boxed{\pi_{i1}(\mathbf{x}_i) = \frac{1}{1 + \sum_{r \neq 1} \exp\left(\eta_{ir}(\mathbf{x}_i)\right)} = \frac{1}{1 + \sum_{r \neq 1} \pi_{ij}(\mathbf{x_i})/\pi_{i1}(\mathbf{x_i})}}$$

POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

5-4.2 Models for ordinal scales

- Ordinal models are very common in practical applications, much more than nominal models.
- In many applications, category definition for the polytomous target variables is subjective. Conclusions have to be consistent to category aggregation among contiguous categories.
- Models are commonly based on cumulative probabilities, $\boxed{\gamma_j = \mathrm{P}(Y \leq j)}$, instead of actual probability of the categories $\boxed{\pi_j = \mathrm{P}(Y = j)}$.
- Both probability sets are equivalent, but cumulative probability models show enhanced properties than actual category models for ordinal scales.
- In particular, GMLz using logit link on cumulative probabilities have proved to work well in practice, $\boxed{\log \gamma_j / (1 - \gamma_j)}$.

POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

- Basic models based on logit transformation of cumulative probabilities:

$$\log \frac{\gamma_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})} = \alpha_j - \beta^T \mathbf{x} \quad j = 1,\ldots,K-1$$

(model A+X)

They are called, **proportional logodd models**, since odds-ratio for $Y \leq j$ given $\mathbf{x} = \mathbf{x}_1$ and $\mathbf{x} = \mathbf{x}_2$ becomes,

$$\frac{\gamma_j(\mathbf{x}_1)/(1-\gamma_j(\mathbf{x}_1))}{\gamma_j(\mathbf{x}_2)/(1-\gamma_j(\mathbf{x}_2))} = e^{-\beta^T(\mathbf{x}_1-\mathbf{x}_2)} \quad j = 1,\ldots,K-1$$

- Negative sign in model parameters β is an convention to guarantee for large $\beta^T \mathbf{x}$ values in the linear predictor high probabilities for last categories.

- α and β parameters are to be estimated subject to the constraint of non-decreasing independent terms, $\boxed{\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{K-1}}$.

POLYTOMOUS TARGET MODELLING: MEASUREMENT SCALES

- Instead of applying *logit link* to cumulative probabilities a *c-log-log link* has been also applied (or log-log complementary link, $g_3(\pi)$), the resulting model is called *proportional-hazards model* ,

$$\log\left(\log\frac{1}{1-\gamma_j(\mathbf{x})}\right)=\alpha_j-\beta^T\mathbf{x}\quad j=1,\ldots,K-1$$

(model X with estimates common to all categories)

- Parameters to be estimated are exactly the same (different values since the underlying scale is the linear predictor is not the same), subject to non-decreasing independent terms, $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{k-1}$, to guarantee non-negative probabilities.

- More complex models considering non-parallel planes in the linear predictor can be formulated, either using *logit link or c-log-log link*:

$$\boxed{\eta_{ij}(\mathbf{x})=\alpha_j-\beta_{\mathbf{j}}^{\mathbf{T}}\mathbf{x}\quad j=1,\ldots,K-1}$$

(model X with estimates depending on categories)

5-5. POLYTOMOUS TARGET MODELLING. LATENT VARIABLE AND DISCRETE CHOICE FORMULATIONS

➡ Random Utility models for discrete choice models are well-known for market analysis and demand models in transportation. They refer to a latent scale, non-observable random variable, called utility.

A latent variable approach, non-observable, is the propension to alternative choice.

5-5.1 Latent variable formulation

➡ Let Y_i be a multinomial variable representing an observable polytomous and discrete response, being 1, 2 to *K*.

➡ Let us assume a non-observable and continuous variable Y_i^*, such that Y_i outcome is connected to intervals in the latent scale Y_i^* according to cut points $\boxed{\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{K-1}}$ in such a way that Y_i takes 1 value whenever $\boxed{Y_i^* < \alpha_1}$. In general, Y_i takes *j* value whenever $\boxed{\alpha_{j-1} < Y_i^* < \alpha_j}$.

➡ Let us assume that Y_i^* follows a linear model $Y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$, where the error term has a probability distribution $F(\varepsilon_i)$.

POLYTOMOUS TARGET. LATENT AND DISCRETE CHOICE FORMULATIONS

➡ Then the probability of the observed outcome for observation i being in actual category less than jth category given explanatory variables in $\mathbf{x_i}$ satisfies the equation:

$$\gamma_{ij} = P\left(Y_i^* < \alpha_j\right) = P\left(\mathbf{x_i^T}\boldsymbol{\beta} + \varepsilon_i < \alpha_j\right) = P\left(\varepsilon_i < \alpha_j - \mathbf{x_i^T}\boldsymbol{\beta}\right) = F\left(\alpha_j - \mathbf{x_i^T}\boldsymbol{\beta}\right)$$

Thus, the relationship between cumulative probability and the linear predictor follows the inverse function of the probability distribution for the latent variable:

$$g\left(\gamma_{ij}\right) = F^{-1}\left(\gamma_{ij}\right) = \alpha_j - \mathbf{x_i^T}\boldsymbol{\beta}$$

➡ A very important consideration when interpreting model estimates is that a simultaneous identification of both $\boldsymbol{\beta}$ and the scale of the probability distribution for the error term is not possible and standardized distributions are considered.

➡ If the error term in the linear model included in the linear predictor for the non-observable latent variable are:

1. $\varepsilon_i \approx N\left(0, \sigma^2\right)$, then probit models hold, and estimated coefficients are $\boldsymbol{\beta}/\sigma$.
2. Error term logistically distributed with standard deviation $\sigma = \pi/\sqrt{3}$, then logit models hold.

POLYTOMOUS TARGET. LATENT AND DISCRETE CHOICE FORMULATIONS

- In the context of transport demand modelling, propension to choice an alternative is defined through unobserved utility maximization:

$$Y_{ij} = \begin{cases} 0 & \\ 1 & \text{Trip maker } i \text{ selects j-th alternative whenever its gives the highest utility among all available } alternatives \end{cases}$$

- Utility is a random variable that can be decomposed into a systematic part and an error term. Systematic part is modelled as a linear combination of explanatory variables and the error term is modelled according to extreme value distributions (Gumbel, Gompertz), normal distributions, etc.
- Whenever error terms are not independent and identically distributed (i.i.d.), generalized extreme value or multinormal distributions are specified.

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x}_{\mathbf{j}} + \varepsilon_{ij}$$

- Probability of choising j-th alternative if maximum probability applies to j-th alternative.

POLYTOMOUS TARGET. LATENT AND DISCRETE CHOICE FORMULATIONS

- Alternatives do not produce utility by themselves, utility is derived from explanatory variables included in the systematic part of the utility model linked to each alternative and some other related to trip maker characteristics.

- Random utility models are individual-based models (disaggregated demand models). They rely on individual behavior and they are technically extensions of generalized linear models.

- R coverts this family of models in the **mlogit package**.

  - Decision maker *i* selects alternative *j* if and only if it exhibits the maximum utility:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \geq U_{il} = V_{il} + \varepsilon_{il} \quad \rightarrow V_{ij} - V_{il} \geq \varepsilon_{il} - \varepsilon_{ij} \;\; \forall l \neq j$$

  - The probability that decision maker *n* chooses alternative *j* is

$$\boxed{P_{ij} = \mathrm{Prob}(U_{ij} > U_{il} \;\forall j <> l) = \mathrm{Prob}(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il} \;\forall j <> l) = \mathrm{Prob}(\varepsilon_{il} - \varepsilon_{ij} < V_{ij} - V_{il} \;\forall j <> l)}$$

  - Using the join density of the random vector $f(\boldsymbol{\varepsilon}_i)$

POLYTOMOUS TARGET. LATENT AND DISCRETE CHOICE FORMULATIONS

- Data on travel mode choice for travel between Sydney and Melbourne, Australia.
  - A data frame containing 840 observations on 4 modes for 210 individuals.
    - A model using specific utility constants and Wait - Terminal waiting time, 0 for car and Gcost - Generalized cost measure.
    - Reference mode is car. See R reports in next pages.

$$V_{air} = 5.77 - 0.097\,TTME_{air} - 0.016(GC_{air})$$
$$V_{train} = 3.92 - 0.097\,TTME_{train} - 0.016(GC_{train})$$
$$V_{bus} = 3.21 - 0.097\,TTME_{bus} - 0.016(GC_{bus})$$
$$V_{car} = 0 - 0.097\,TTME_{car} - 0.016(GC_{car})$$

$$\pi_{car} = \frac{\exp(V_{car})}{\exp(V_{air}) + \exp(V_{train}) + \exp(V_{bus}) + \exp(V_{car})}$$

5-6. POLYTOMOUS TARGETS: HIERARCHICAL APPROACH

This is an straightforward extension of models for binary targets. A set of dichotomic decision models.

➡ For exemple, *K=4*

1 2		3 4	
↓		↓	
1	2	3	4

1	2 3 4	
	↓	
	2	3 4
		↓
	3	4

➡ ***A hierarchical set of binary decisions is a very natural approach whenever decisions represent an ordered decision process.*** For example, in a modal choice decision process in transport modelling for Barcelona Metropolitan Area.

Private		**Public Transport**	
↓		↓	
Auto	**Taxi**	**Bus**	**Fixed Infrastructure**
			↓
		Metro	**Train**

➡ **It is not easy to identify if a hierarchical decision process is suitable for modal choice modelling.**

5-7. POLYTOMOUS TARGETS. EXAMPLES.

5-7.1 Womenlf by Fox

A generalized linear model can be proposed to analyse the relationship between married women working status (polytomous target) and explanatory variables Children presence (binary factor, Yes or No) and Income group (5 levels, polytomous factor) and residence region (Region).

- Target response **variable is polytomous with** 3 categories: *doesn't work (1), partial work (2) and fulltime work (3)*. Baseline category is '*doesn't work*'.
- **Factor A**, Children, with 2 categories (Yes, No). Baseline category: No (constant corresponds to the average linear predictor effect for No category.
- **Factor B**, Canada Region, has 5 categories.
- **Factor C**, Income group (in thousands of Canadian dollars) can be also treated as a numeric covariate X.
- Intuition indicates an interaction between Income and Children presence. A*X.

POLYTOMOUS TARGETS. EXAMPLES.

```
WOMEN'S LABOUR-FORCE PARTICIPATION DATASET, CANADA 1977

[1] OBSERVATION
[2] LABOUR-FORCE PARTICIPATION
     fulltime = WORKING FULL-TIME
     parttime = WORKING PART-TIME
     not_work = NOT WORKING   OUTSIDE THE HOME
[3] HUSBAND'S IINCOME, $1000'S

[4] PRESENCE OF CHILDREN
     absent
     present
[5] REGION
     Atlantic = ATLANTIC CANADA
     Quebec
     Ontario
     Prairie  = PRAIRIE PROVINCES
     BC       = BRITISH COLUMBIA
Source: Social Change in Canada Project, York Institute for Social Research.
DATA:
  1         not_work        15         present         Ontario
  2         not_work        13         present         Ontario
…
253         not_work        13         present         Quebec
254         parttime        23         present         Quebec
255         fulltime        11         absent          Quebec
…
263         not_work        15         present         Quebec
ENDDATA
```

POLYTOMOUS TARGETS. EXAMPLES.

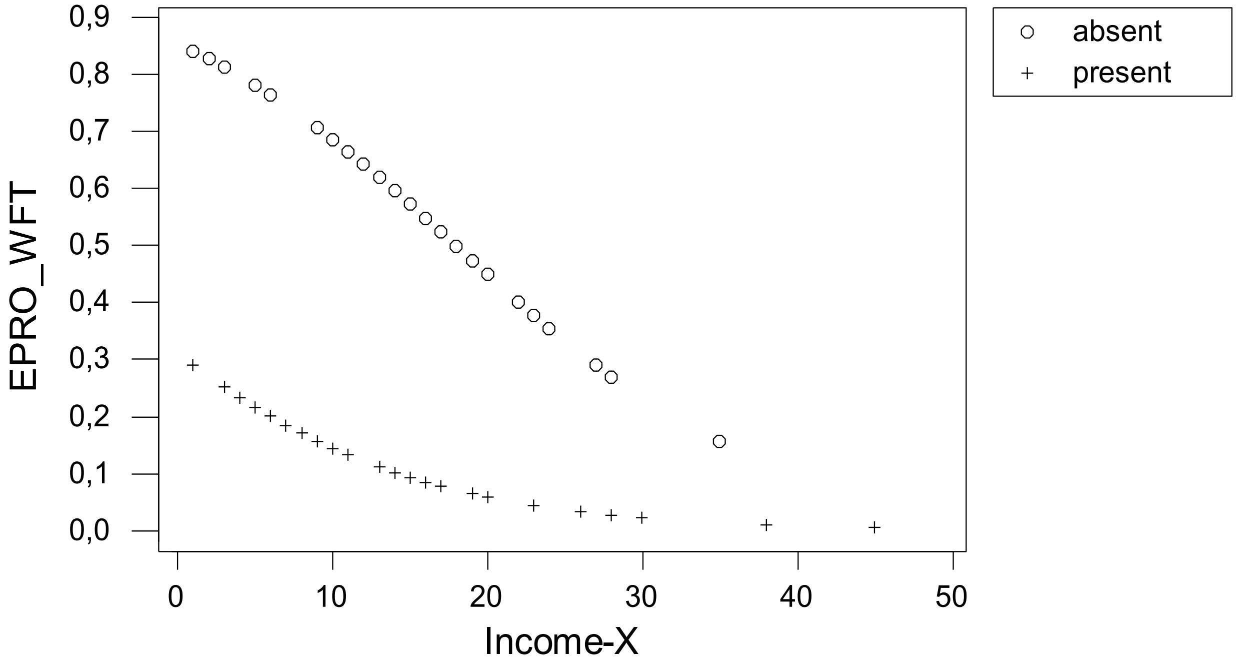
Deviance table is presented for several models. The most suitable model contains X and A, whose negative coefficient indicates that whenever children are present and Income decreases women fulltime work.

		Deviance Analysis					
Model		p	~~Deviance or~~ Log-likelihood	ΔDevianza	d.f.	Comments: Contrast	Comments: H_0 Accept.
0	1	2	¿?	86.439	14	0 vs 8	No
1	A	4	-219.018	15.154	2	1 vs 3	No
2	X	4	-243.220	63.558	2	2 vs 3	No
3	A+X	6	-211.441	7.416	8	3 vs 7	Yes
4	A+B	12	-215.055	14.644	2	4 vs 7	No
5	B+X	12	-240.335	65.204	2	5 vs 7	No
6	A*X	8	-210.715	7.286	8	6 vs 8	Yes
7	A+B+X	14	-207.733	1.322	2	7 vs 8	Yes
8	B+A*X	16	-207.072				$\chi^2_{2,0.05} = 5.991$

POLYTOMOUS TARGETS. EXAMPLES.

$$\log \frac{\pi_{i3}}{\pi_{i1}} = 1.983 - 2.559\, Factor\ A_i - 0.09723 x_i$$

where $Factor\ A_i = 1$ if children are present and 0 otherwise.



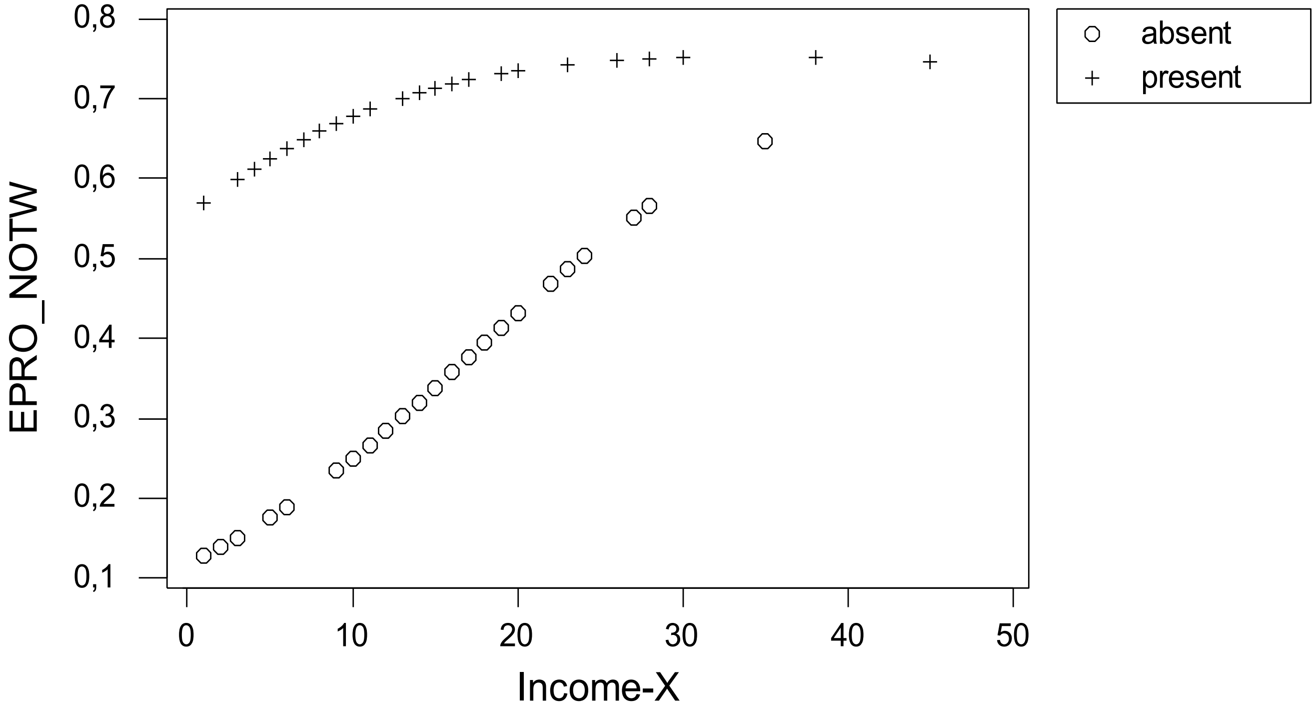
➡ M7 vs M8 contrast indicates that husband income interaction with children presence (Factor A) is not statistically significance.

➡ M3 vs M7 contrast indicates that región (Factor B) is not statistically significance.

➡ Nevertheless, Factor A (M1 vs M3) and covariate X (M2 vs M3) main gross effects are statistically significant (corresponding null hypothesis are rejected).

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.



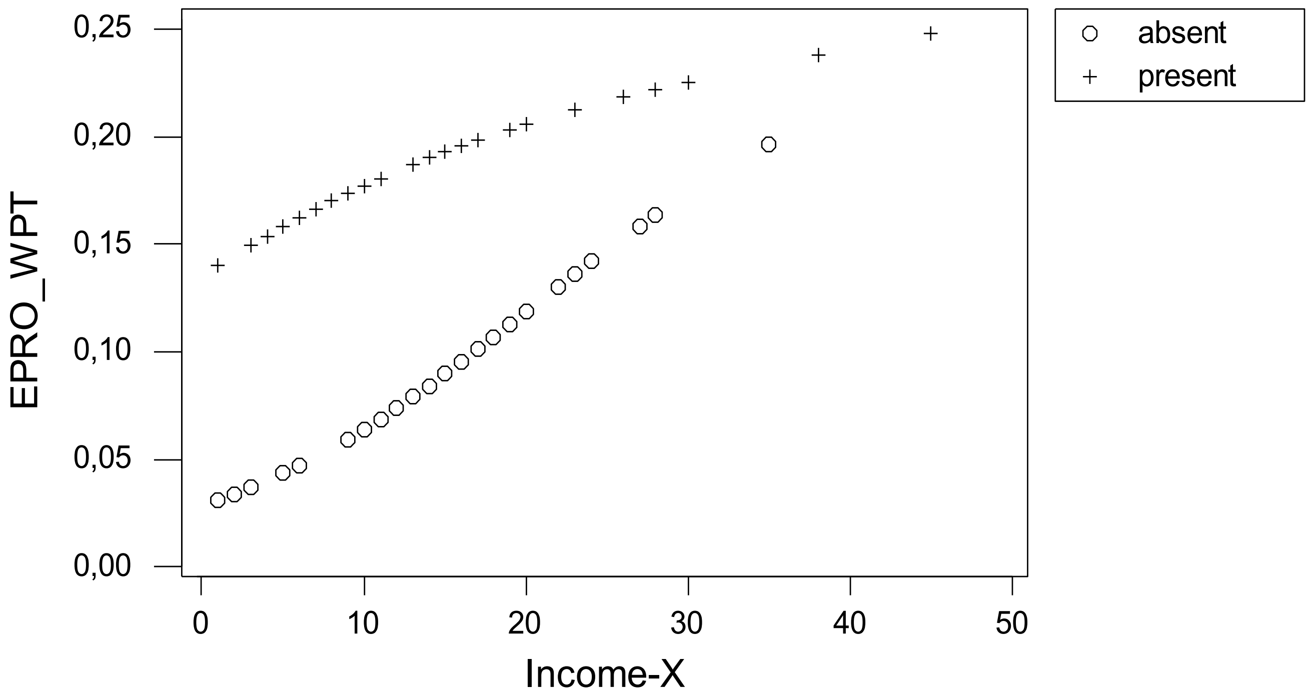
Increasing husband income and having kids at home, don't work category increases.

Nevertheless, partial time work seems almost unaffected

POLYTOMOUS TARGETS. EXAMPLES.

$$\log \frac{\pi_{i2}}{\pi_{i1}} = -1.43 + 0.022\,Factor\ A_i + 0.0069 x_i$$

where $Factor\ A_i = 1$ if children are present and 0 otherwise.



Residual analysis vs estimated probabilities can not be directly done.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

```
MTB > Name c7 = 'NTRI1'
MTB > NLogistic 'Y_i' = 'Factor A' 'Income-X';
SUBC>   Factors 'Factor A';
SUBC>   Reference 'Y_i' 'not_work';
SUBC>   Ntrials 'NTRI1';
SUBC>   Brief 3.
```

Nominal Logistic Regression: Y_i versus Factor A; Income-X

Response Information

Variable	Value	Count	
Y_i	not_work	155	(Reference Event)
	parttime	42	
	fulltime	66	
	Total	263	

Factor Information

Factor	Levels	Values
Factor A	2	absent present

Logistic Regression Table

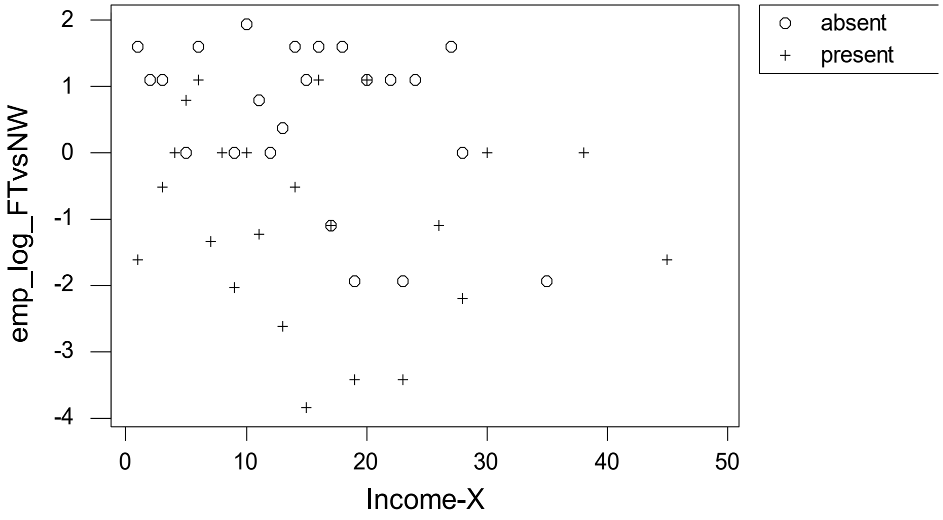
Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Logit 1: (parttime/not_work)							
Constant	-1,4323	0,5925	-2,42	0,016			
Factor A							
present	0,0215	0,4690	0,05	0,963	1,02	0,41	2,56
Income-X	0,00689	0,02345	0,29	0,769	1,01	0,96	1,05
Logit 2: (fulltime/not_work)							
Constant	1,9828	0,4842	4,10	0,000			
Factor A							
present	-2,5586	0,3622	-7,06	0,000	0,08	0,04	0,16
Income-X	-0,09723	0,02810	-3,46	0,001	0,91	0,86	0,96

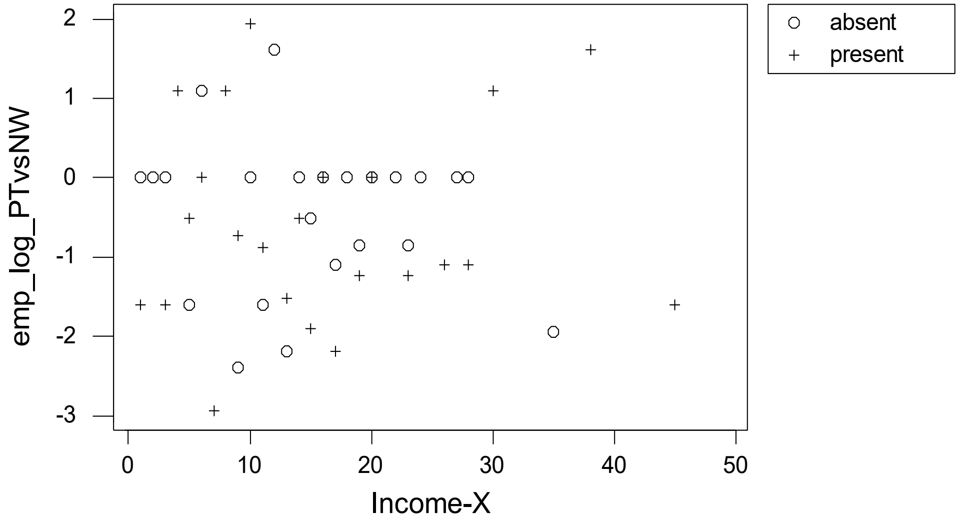
POLYTOMOUS TARGETS. EXAMPLES.

```
Log-likelihood = -211,441
Test that all slopes are zero: G = 77,611; DF = 4; P-Value = 0,000

Goodness-of-Fit Tests

Method        Chi-Square    DF      P
Pearson          164,769    86  0,000
Deviance         138,674    86  0,000

MTB >
```

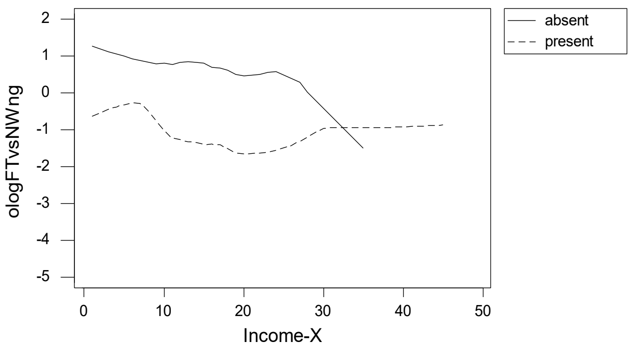
➡ Logodd empirical transformation vs predicted logodds by the model have to be compared. $\log\left(\frac{y_{ij}+\frac{1}{2}}{y_{i1}+\frac{1}{2}}\right)$ and linear relationship has to be validated setting 1st category as baseline.

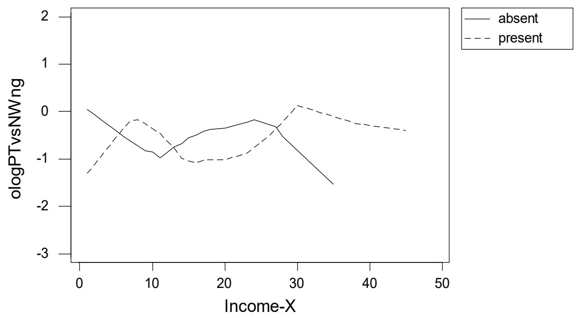
POLYTOMOUS TARGETS. EXAMPLES.





➡ Observed log-odds 'Full time' and 'Partial time' vs 'Does Not work'.

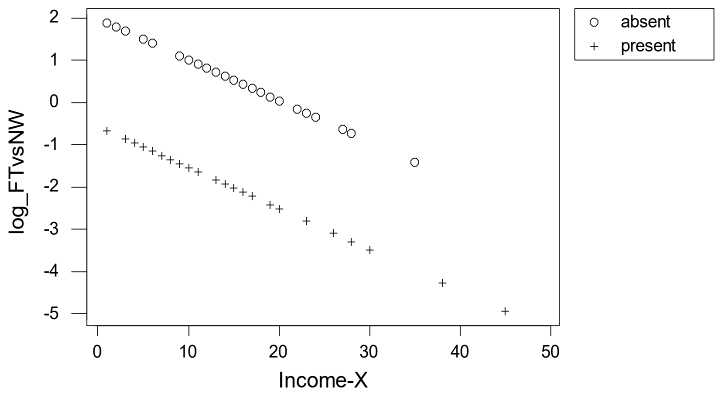


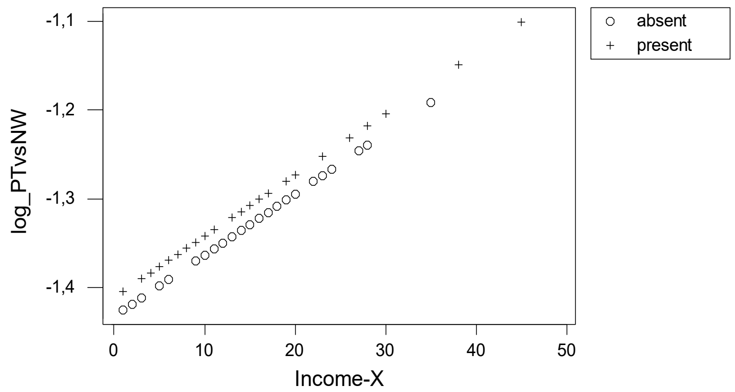


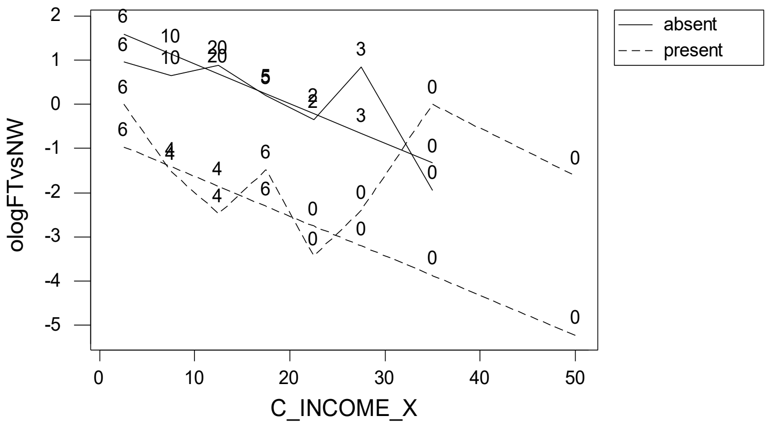
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

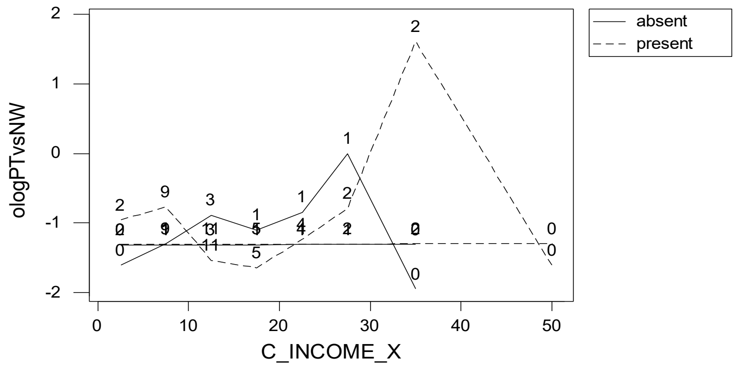
POLYTOMOUS TARGETS. EXAMPLES.

➡ Model A+X









POLYTOMOUS TARGETS. EXAMPLES. R OUTPUT

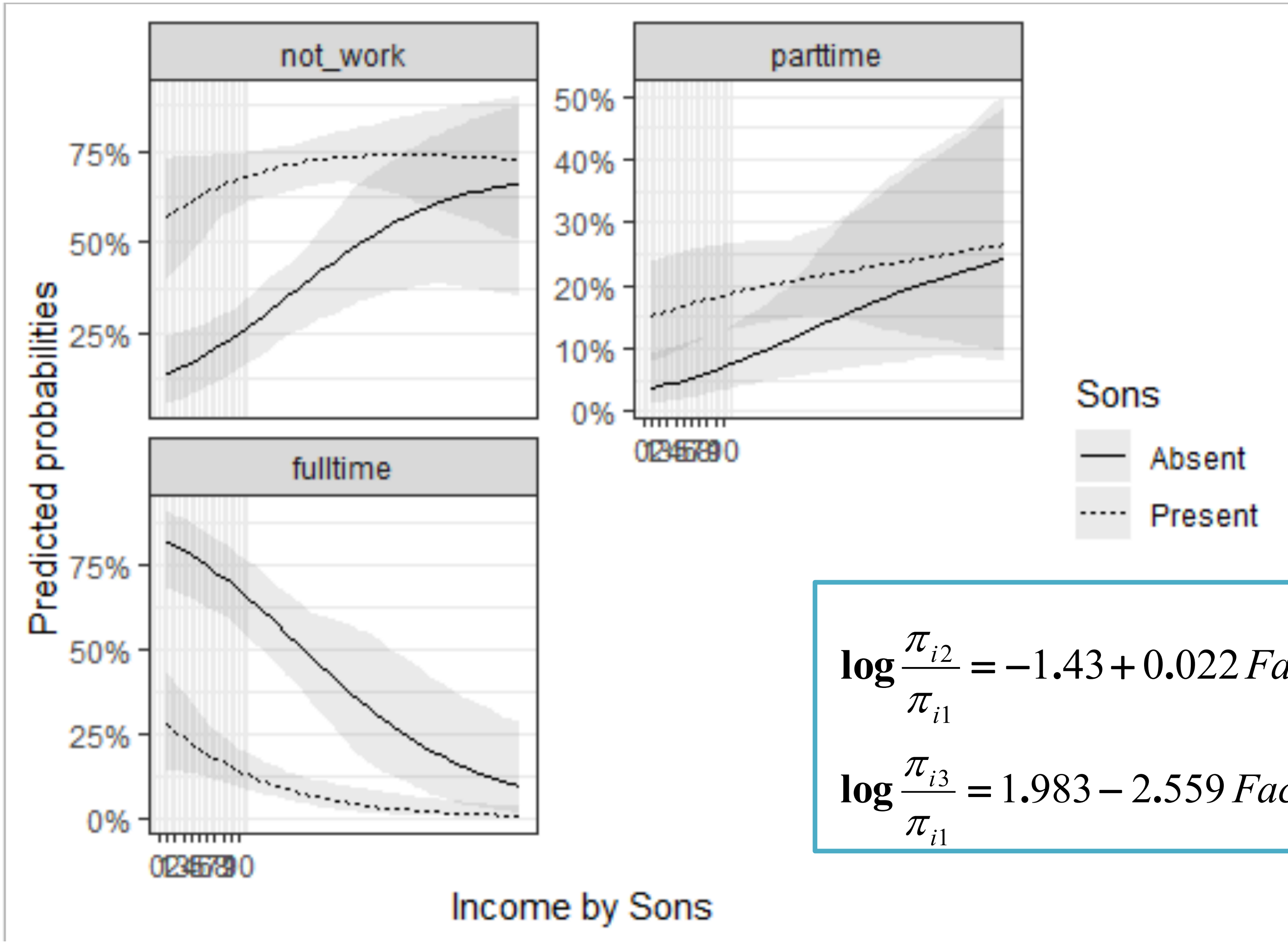
```
> summary(mm3)
Call: multinom(formula = work ~ sons + income, data = womenlf, weights = ones)

Coefficients:
         (Intercept) sonspresent       income
parttime   -1.432321  0.02145558  0.006893838
fulltime    1.982842 -2.55860537 -0.097232073

Std. Errors:
         (Intercept) sonspresent     income
parttime   0.5924627   0.4690352 0.02345484
fulltime   0.4841789   0.3621999 0.02809599

Residual Deviance: 422.8819
AIC: 434.8819

> predict(mm3, type="probs",newdata=data.frame(sons="present",income=14.75665))
 not_work  parttime  fulltime
0.7122650 0.1923548 0.0953802
```

$$\log \frac{\pi_{i2}}{\pi_{i1}} = -1.43 + 0.022\, Factor\ A_i + 0.0069 x_i$$

$$\log \frac{\pi_{i3}}{\pi_{i1}} = 1.983 - 2.559\, Factor\ A_i - 0.09723 x_i$$

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES. R OUTPUT



$$\log\frac{\pi_{i2}}{\pi_{i1}} = -1.43 + 0.022\,Factor\ A_i + 0.0069x_i$$

$$\log\frac{\pi_{i3}}{\pi_{i1}} = 1.983 - 2.559\,Factor\ A_i - 0.09723x_i$$

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

➡ Ordinal scale proposal. Resulting models

$$\log \frac{\gamma_{i1}}{1-\gamma_{i1}} = \log \frac{\pi_{i1}}{\pi_{i2}+\pi_{i3}} = -1.852 + 1.972\, Factor\ A_i + 0.0539\, x_i$$

$$\log \frac{\gamma_{i2}}{1-\gamma_{i2}} = \log \frac{\pi_{i1}+\pi_{i2}}{\pi_{i3}} = -0.9409 + 1.972\, Factor\ A_i + 0.0539\, x_i$$

where $Factor\ A_i = 1$ if children are present and 0 otherwise.

```
MTB > OLogistic 'Y_i' = 'Factor A' 'Income-X';
SUBC>   Factors 'Factor A';
SUBC>   Logit;
SUBC>   Order 'not_work' 'parttime' 'fulltime' ;
SUBC>   Brief 3.
Ordinal Logistic Regression: Y_i versus Factor A; Income-X
Link Function:  Logit
Response Information
Variable  Value        Count
Y_i       not_work       155
          parttime        42
          fulltime        66
          Total          263
```

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

```
Factor Information

Factor   Levels Values
Factor A      2 absent  present

Logistic Regression Table
                                                     Odds       95% CI
Predictor        Coef     SE Coef        Z      P    Ratio    Lower    Upper
Const(1)      -1,8520      0,3773    -4,91  0,000
Const(2)      -0,9409      0,3619    -2,60  0,009
Factor A
 present       1,9720      0,2804     7,03  0,000     7,18     4,15    12,45
Income-X      0,05390     0,01943     2,77  0,006     1,06     1,02     1,10

Log-likelihood = -220,831
Test that all slopes are zero: G = 58,830; DF = 2; P-Value = 0,000

Goodness-of-Fit Tests

Method      Chi-Square    DF      P
Pearson        175,341    88  0,000
Deviance       157,455    88  0,000

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs            Number  Percent     Summary Measures
Concordant        13800    70,7%     Somers' D                0,45
Discordant         5090    26,1%     Goodman-Kruskal Gamma    0,46
Ties                622     3,2%     Kendall's Tau-a          0,25
Total             19512   100,0%
```

POLYTOMOUS TARGETS. EXAMPLES.

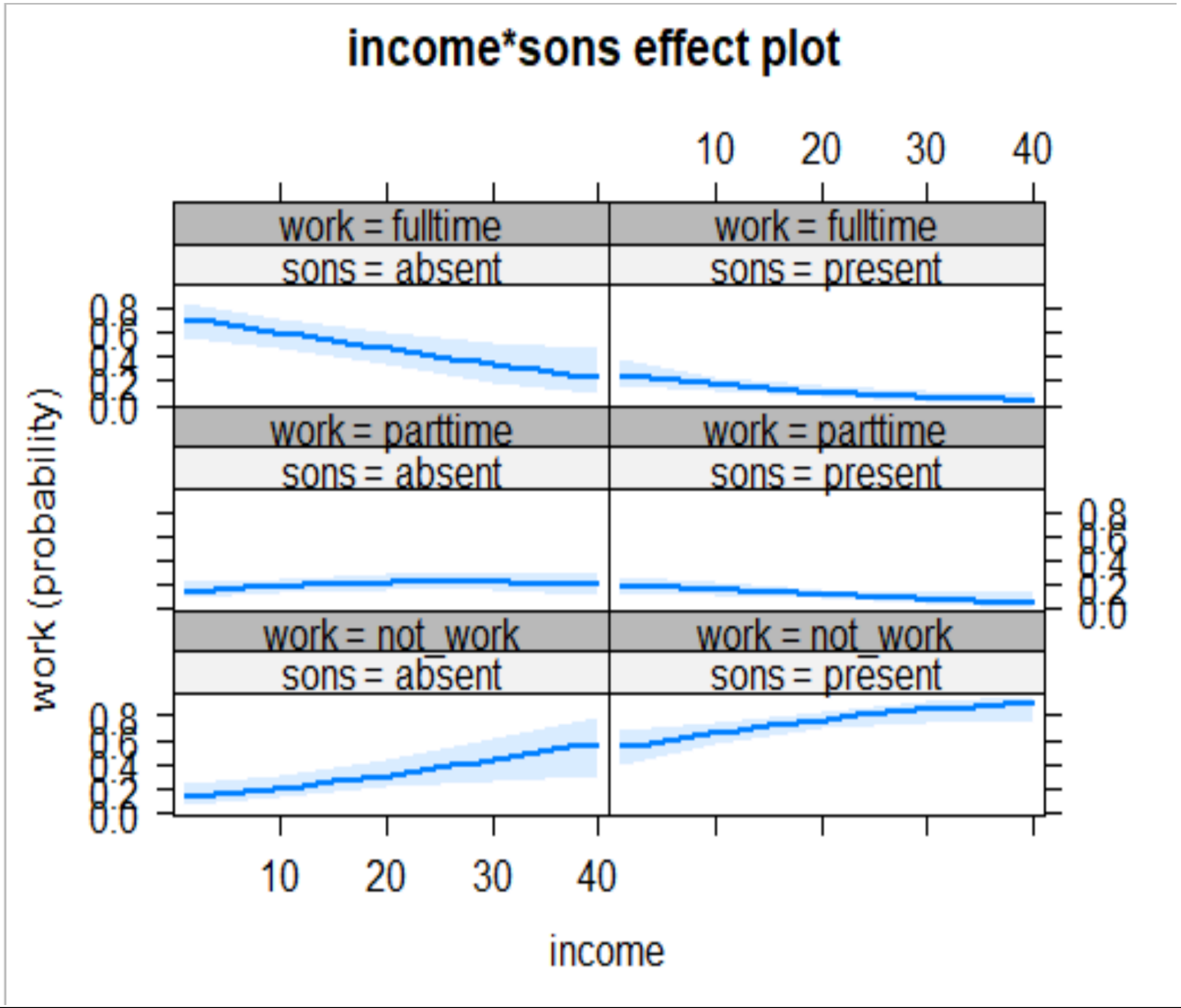
- Ordinal treatment using R: polr() method that operates under latent variable paradigm (sign to be changed).

```
> summary(womenlf[,1:6])
       id               work          income            sons            region        ones  
 Min.   :  1.0   not_work:155   Min.   : 1.00   absent : 79   Atlantic: 30   Min.   :1  
 1st Qu.: 66.5   parttime: 42   1st Qu.:10.00   present:184   BC      : 29   1st Qu.:1  
 Median :132.0   fulltime: 66   Median :14.00                 Ontario :108   Median :1  
 Mean   :132.0                  Mean   :14.76                 Prairie : 31   Mean   :1  
 3rd Qu.:197.5                  3rd Qu.:19.00                 Quebec  : 65   3rd Qu.:1  
 Max.   :263.0                  Max.   :45.00                                Max.   :1  

> om3 <- polr(work~income+sons,data=womenlf,weight=ones)
> summary(om3)
Call:
polr(formula = work ~ income + sons, data = womenlf, weights = ones)
Coefficients:
              Value Std. Error t value
income      -0.0539    0.01949  -2.766
sonspresent -1.9720    0.28695  -6.872

Intercepts:
                  Value   Std. Error t value
not_work|parttime -1.8520  0.3863    -4.7943
parttime|fulltime -0.9409  0.3699    -2.5435
Residual Deviance: 441.663
AIC: 449.663
```

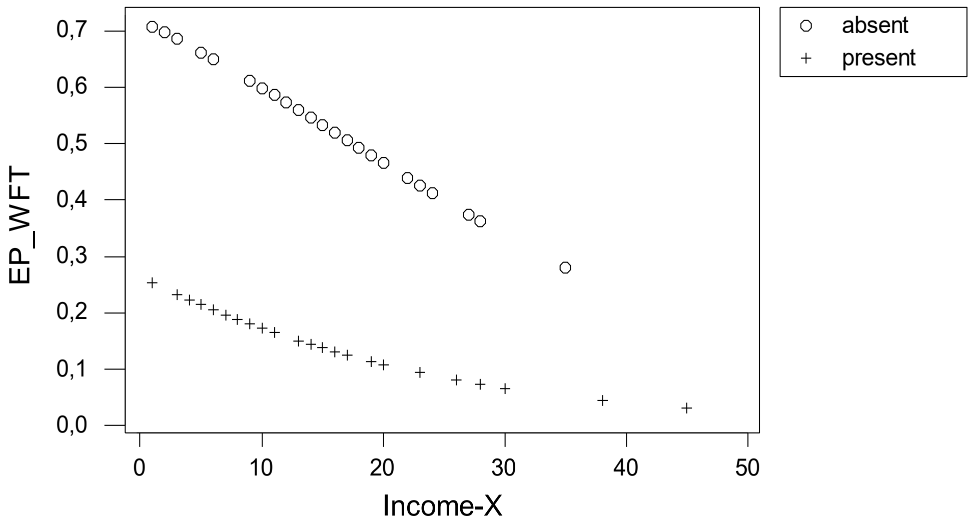
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

```
> Effect(focal.predictors = c("income","sons"), om3)
> plot(Effect(focal.predictors = c("income","sons"), om3))
```



POLYTOMOUS TARGETS. EXAMPLE: MINITAB USING REVERSE ORDER



➡ Polytomous target categories can be ordered as : *doesn't work (1), parttime work (2) and fulltime work (3).*

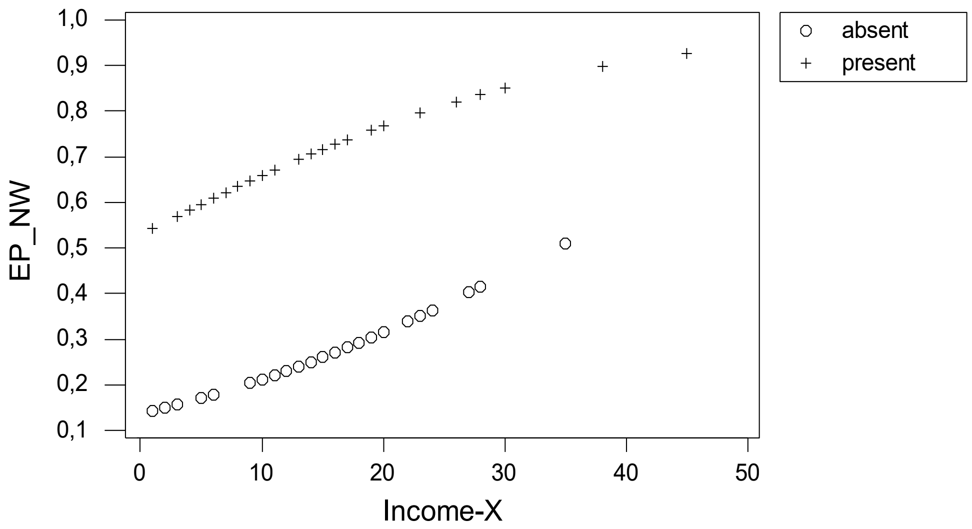
➡ Resulting ordinal model estimates:

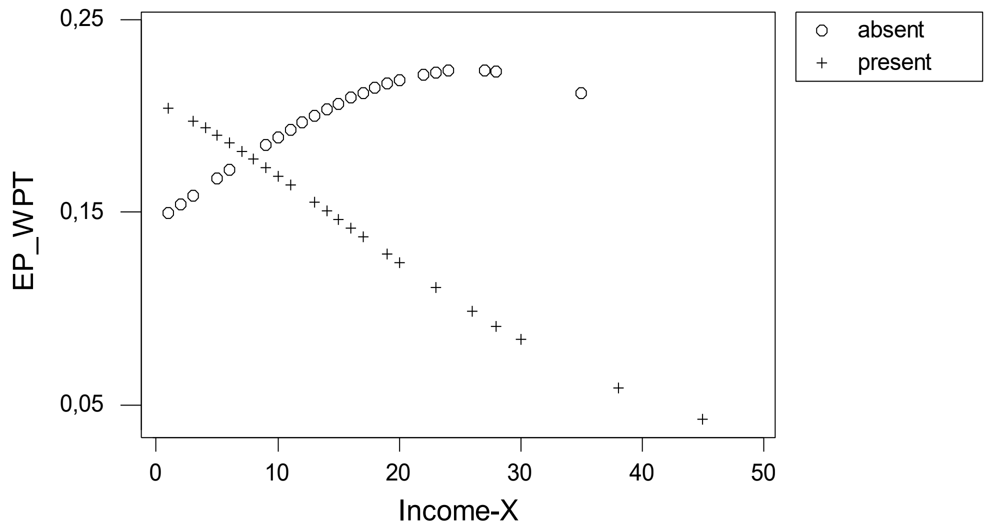
If $\boxed{Factor\ A_i = 1}$ children are present and 0 otherwise

$$\log\frac{\gamma_{i1}}{1-\gamma_{i1}} = \log\frac{\pi_{i1}}{\pi_{i2}+\pi_{i3}} = -1.852 + 1.972\,Factor\ A_i + 0.0539\ x_i$$

$$\log\frac{\gamma_{i2}}{1-\gamma_{i2}} = \log\frac{\pi_{i1}+\pi_{i2}}{\pi_{i3}} = -0.941 + 1.972\,Factor\ A_i + 0.0539\,x_i$$

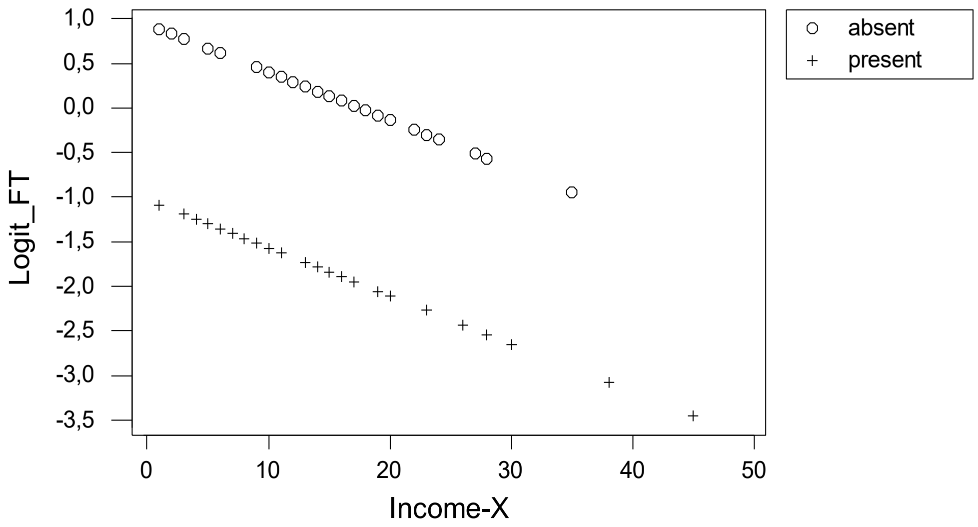
POLYTOMOUS TARGETS. EXAMPLES.

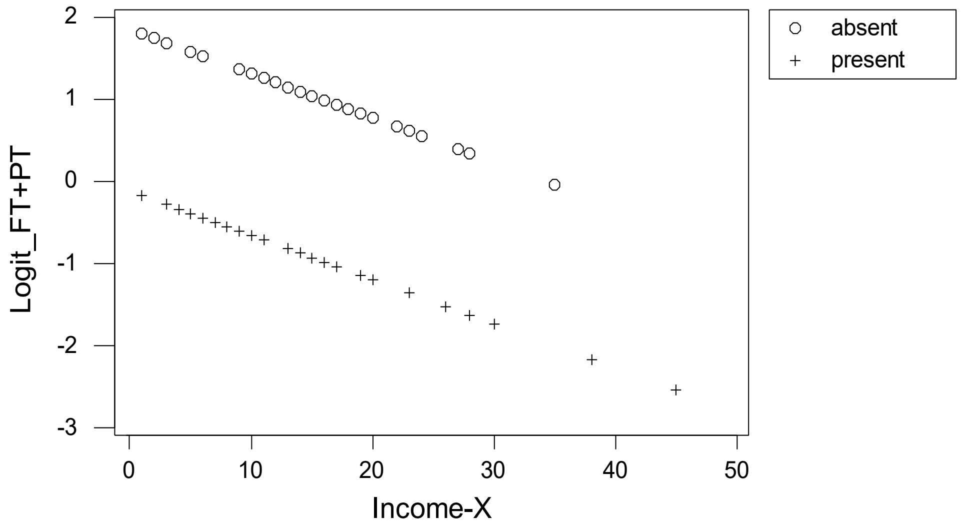




POLYTOMOUS TARGETS. EXAMPLES.

➡ Model predicted: linear predictor scale





POLYTOMOUS TARGETS. EXAMPLES.

5-7.1.1 Hierarchical modelling proposal

Hierarchical proposal
First level: *Not Work (1 - Reference) vs Work (2 and 3)*
Second level: Parttime (2-Reference) vs Fulltime (3) on the subset of those categories belonging to 2 and 3

First level, binary model,

$$\log \frac{\pi_{i2} + \pi_{i3}}{\pi_{i1}} = 1.336 - 1.576\, Factor\ A_i - 0.04231 x_i$$

where $Factor\ A_i = 1$ if children are present and 0 otherwise.

```
> summary(h1bm3)

Call: glm(formula = bwork ~ income + sons, family = binomial, data = womenlf)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.33583    0.38376   3.481   0.0005 ***
income      -0.04231    0.01978  -2.139   0.0324 *
sonspresent -1.57565    0.29226  -5.391    7e-08 ***
---

    Null deviance: 356.15  on 262  degrees of freedom
Residual deviance: 319.73  on 260  degrees of freedom
AIC: 325.73
```

POLYTOMOUS TARGETS. EXAMPLES.

```
Binary Logistic Regression: Ybin_i versus Factor A; Income-X
Step  Log-Likelihood
   0        -178,075
   1        -159,965
   2        -159,866
   3        -159,866
   4        -159,866

Link Function:  Logit
Response Information
Variable  Value          Count
Ybin_i    work             108  (Event)
          not_work         155
          Total            263

Logistic Regression Table
                                                   Odds       95% CI
Predictor       Coef    SE Coef        Z      P    Ratio    Lower    Upper
Constant      1,3358     0,3838     3,48  0,000
Factor A
 present     -1,5756     0,2923    -5,39  0,000     0,21     0,12     0,37
Income-X    -0,04231    0,01978    -2,14  0,032     0,96     0,92     1,00

Log-Likelihood = -159,866
Test that all slopes are zero: G = 36,418; DF = 2; P-Value = 0,000

Goodness-of-Fit Tests
Method                Chi-Square    DF      P
Pearson                   73,229    43  0,003
Deviance                  78,469    43  0,001
Hosmer-Lemeshow            5,824     7  0,560
```

POLYTOMOUS TARGETS. EXAMPLES.

Second level: *Parttime* (*baseline 2*) and *Fulltime* (3):

$$\log \frac{\pi_{i3}}{\pi_{i2}} = 3.478 - 2.651\, Factor\ A_i - 0.1073\, x_i$$

where $Factor\ A_i = 1$ if children are present and 0 otherwise.

```
> summary(h2bm3)

Call: glm(formula = work ~ income + sons, family = binomial, data = womenlf,subset = bwork ==
"work")

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.47777    0.76711   4.534 5.80e-06 ***
income      -0.10727    0.03915  -2.740  0.00615 **
sonspresent -2.65146    0.54108  -4.900 9.57e-07 ***
---
    Null deviance: 144.34  on 107  degrees of freedom
Residual deviance: 104.49  on 105  degrees of freedom
AIC: 110.49
```

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

➡ **How to select the best proposal?**

➡ Interesting results are obtained: children factor A and Income-X effects are enforced in logodds for fulltime work than parttime work in the hierarchical proposal at the second level. First level effects for the children factor A and X Income covariate are not so intense in the logodds *Work vs NotWork*.

➡ Hierarchical proposal does not behave consistent to nominal proposal.

Best Nominal, hierarchical and ordinal proposals can be compared by using AIC/BIC statistics. Minimum AIC corresponds to the NOMINAL multinomial modelling.

- AIC Nominal: 2(211,441+6) = 434.882
- AIC Hierarchical: Sum AIC from the two levels (take log-likelihood): 2(159.866+3) + 2(52.247+3)=436.266
- AIC Ordinal: 2(220.831 + 4) = 449.662

POLYTOMOUS TARGETS. EXAMPLES.

5-7.2 Cheese tasting (McCullagh)

➡ Ordinal target with 9 categories 1 a 9 meaning *extremely bad* (1) to excellent taste (9) depending on 4 additives, A B C and D (baseline).

	Target variable (Y)									
Additives	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
A	*0*	*0*	*1*	*7*	*8*	*8*	*19*	*8*	*1*	52
B	*6*	*9*	*12*	*11*	*7*	*6*	*1*	*0*	*0*	52
C	*1*	*1*	*6*	*8*	*23*	*7*	*5*	*1*	*0*	52
D	*0*	*0*	*0*	*1*	*3*	*7*	*14*	*16*	*11*	52
Total	7	10	19	27	41	28	39	25	12	208

```
MTB > Name c4 = 'NTRI1' c5 = 'EPROB1' c6 = 'EPROB2' c7 = 'EPROB3' &
CONT>       c8 = 'EPROB4' c9 = 'EPROB5' c10 = 'EPROB6' c11 = 'EPROB7' &
CONT>       c12 = 'EPROB8' c13 = 'EPROB9' c14 = 'CUMP1' c15 = 'CUMP2' &
CONT>       c16 = 'CUMP3' c17 = 'CUMP4' c18 = 'CUMP5' c19 = 'CUMP6' &
CONT>       c20 = 'CUMP7' c21 = 'CUMP8' c22 = 'NOCC1' c23 = 'NOCC2' &
CONT>       c24 = 'NOCC3' c25 = 'NOCC4' c26 = 'NOCC5' c27 = 'NOCC6' &
CONT>       c28 = 'NOCC7' c29 = 'NOCC8' c30 = 'NOCC9'
```

POLYTOMOUS TARGETS. EXAMPLES.

```
MTB > OLogistic 'Response' = 'Factor A';
SUBC>   Frequency 'Y_ij';
SUBC>   Factors 'Factor A';
SUBC>   Logit;
SUBC>   Order 1 2 3 4 5 6 7 8 9;
SUBC>   Reference  'Factor A' 'D';
SUBC>   Ntrials 'NTRI1';
SUBC>   Eprobability 'EPROB1'-'EPROB9';
SUBC>   Cumprobability 'CUMP1'-'CUMP8';
SUBC>   Noccur 'NOCC1'-'NOCC9';
SUBC>   Brief 3.
Ordinal Logistic Regression: Response versus Factor A
Link Function:  Logit
Response Information
Variable  Value       Count
Response  1               7
          2              10
          3              19
          4              27
          5              41
          6              28
          7              39
          8              25
          9              12
          Total         208
Frequency:  Y_ij
Factor Information
Factor   Levels Values
Factor A       4 D A B C
   28 cases were used
    8 cases contained missing values or was a case with zero frequency.
```

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-7,0802	0,5624	-12,59	0,000			
Const(2)	-6,0250	0,4755	-12,67	0,000			
Const(3)	-4,9254	0,4272	-11,53	0,000			
Const(4)	-3,8568	0,3902	-9,88	0,000			
Const(5)	-2,5206	0,3431	-7,35	0,000			
Const(6)	-1,5685	0,3086	-5,08	0,000			
Const(7)	-0,0669	0,2658	-0,25	0,801			
Const(8)	1,4930	0,3310	4,51	0,000			
Factor A							
A	1,6128	0,3778	4,27	0,000	5,02	2,39	10,52
B	4,9646	0,4741	10,47	0,000	143,26	56,56	362,83
C	3,3227	0,4251	7,82	0,000	27,73	12,06	63,81

Tests for terms with more than 1 degree of freedom

Term	Chi-Square	DF	P
Factor A	115,153	3	0,000

Log-likelihood = -355,674
Test that all slopes are zero: G = 148,454; DF = 3; P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	20,938	21	0,463
Deviance	20,308	21	0,502

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLES.

```
Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs              Number  Percent     Summary Measures
Concordant          12602    67,6%     Somers' D                  0,58
Discordant           1830     9,8%     Goodman-Kruskal Gamma      0,75
Ties                 4203    22,6%     Kendall's Tau-a            0,50
Total               18635   100,0%

MTB > Save  "G:\LIDIA\MLGz2000\MLGZ_00_1\Poli_no_ex4.mpj";
SUBC>    Project;
SUBC>    Replace.
Saving file as: G:\LIDIA\MLGz2000\MLGZ_00_1\Poli_no_ex4.mpj
* NOTE  * Existing file replaced.
MTB >
```

By visual inspection, order of preference according to additives: D, A, C and B.

Ordinal model formulation:

$$\log \frac{\gamma_{ij}}{1-\gamma_{ij}} = \alpha_j + \beta_i$$

1. 8 parameters related to independent terms from 1 to 8. Ordered constant estimates can be seen.
2. 3 extra parameters related to dummy variables for additive type, being baseline level 4, $\hat{\beta}_4 = 0$.

POLYTOMOUS TARGETS. EXAMPLES.

- Dummy variable estimates are consistent to visual inspection D, A, C and B.
- Deviance reduction for additive model with respect to null model: $\beta = \mathbf{0}$ is de 148.45 units that has to be contrasted against a chi squared distribution with 3 d.f (24-21).
- Residual deviance is 20.31 and 21 degrees of freedom are left (d.f.). A goodness of fit test stated as 'H0: Current model is consistent to data' is asymptotically distributed as a chi squared (21 d.f.). Assymptotical approximation is not good because there are few observations in some cells of the table.
- Pearson residuals are under 2.3 units in absolute value.
- Model fit assessment by comparing observed logodds vs predicted logodds has to be addressed.

POLYTOMOUS TARGETS. EXAMPLE IN R.

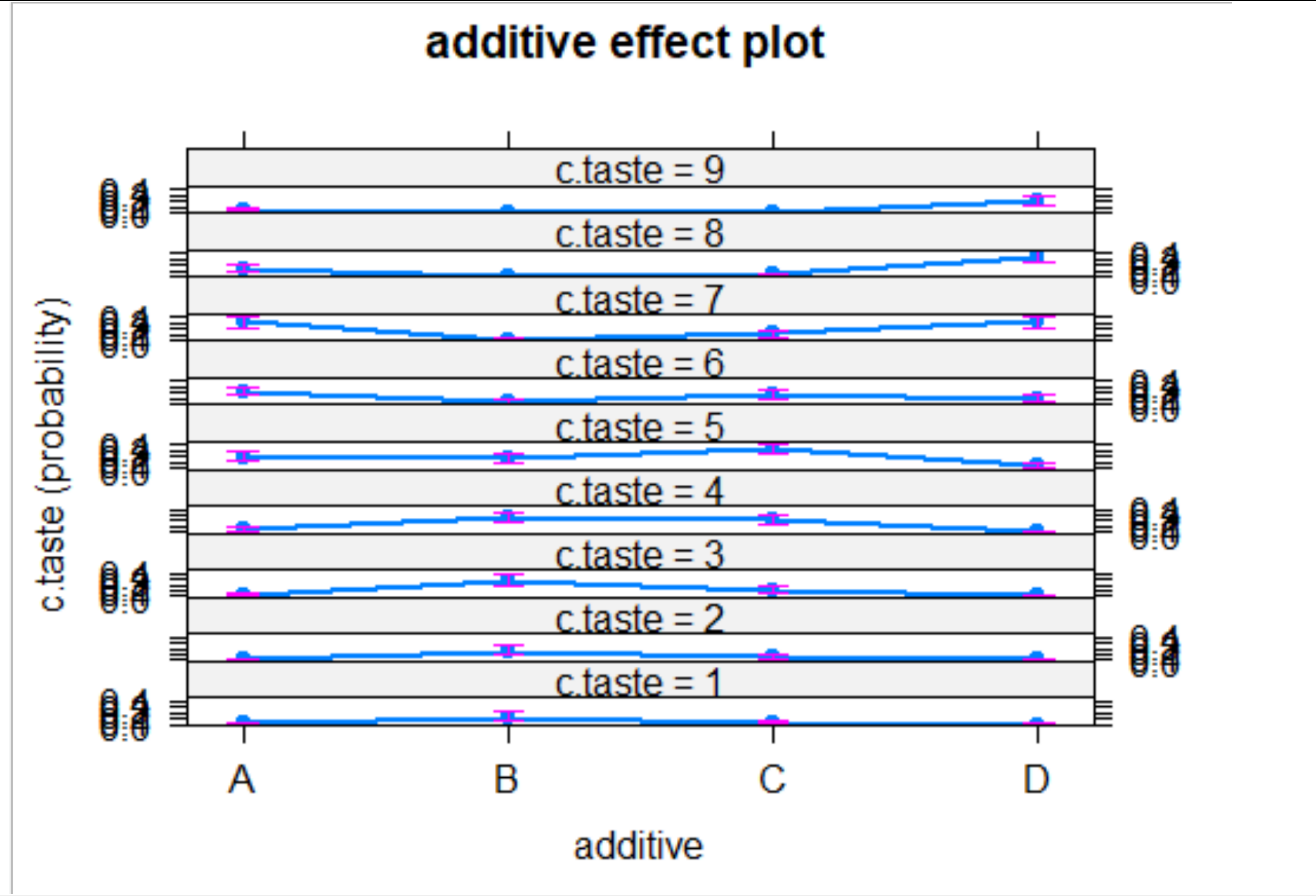
```
> om0 <- polr( c.taste ~ 1, data = cheese, weights=y )
> om1 <- polr( c.taste ~ additive, data = cheese, weights=y )
> om1p <- polr( c.taste ~ additive, data = cheese, method="probit",weights=y )
> om1cg <- polr( c.taste ~ additive, data = cheese, method="cloglog",weights=y )
> anova( om0 , om1, test="Chisq")
Likelihood ratio tests of ordinal regression models
Response: c.taste
     Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
1        1       200   859.8018
2 additive       197   711.3479 1 vs 2     3 148.4539       0
> summary(om1)
Call: polr(formula = c.taste ~ additive, data = cheese, weights = y)

Coefficients:
           Value Std. Error t value
additiveB -3.352     0.4287  -7.819
additiveC -1.710     0.3715  -4.603
additiveD  1.613     0.3805   4.238

Intercepts:
    Value    Std. Error t value
1|2  -5.4674   0.5236   -10.4413
2|3  -4.4122   0.4278   -10.3148
3|4  -3.3126   0.3700    -8.9522
4|5  -2.2440   0.3267    -6.8680
5|6  -0.9078   0.2833    -3.2037
6|7   0.0443   0.2646     0.1673
7|8   1.5459   0.3017     5.1244
8|9   3.1058   0.4057     7.6547
Residual Deviance: 711.3479
AIC: 733.3479
> summary(om1p)
```

$$log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \alpha_j - \beta_i \;\; i = 1 \equiv A\;(ref)$$

or

$$probit(\gamma_{ij}) = \alpha_j - \beta_i \;\; i = 1 \equiv A\;(ref)$$

```
Call:polr(formula = c.taste ~ additive, data = cheese, weights = y, method = "probit")
Coefficients:
            Value Std. Error t value
additiveB -1.8976     0.2267  -8.371
additiveC -0.9766     0.2090  -4.672
additiveD  0.9642     0.2116   4.556

Intercepts:
    Value    Std. Error t value
1|2  -3.1119   0.2661   -11.6936
2|3  -2.5444   0.2260   -11.2562
3|4  -1.8986   0.1948    -9.7447
4|5  -1.2714   0.1744    -7.2902
5|6  -0.4999   0.1605    -3.1139
6|7   0.0488   0.1553     0.3139
7|8   0.9366   0.1690     5.5411
8|9   1.8422   0.2178     8.4564
> om1p$coef
 additiveB  additiveC  additiveD
-1.8975973 -0.9765584  0.9642434
> om1p$zeta - om1p$coef[3]
        1|2         2|3         3|4         4|5         5|6         6|7         7|8
-4.07618241 -3.50868877 -2.86280151 -2.23563544 -1.46413651 -0.91548434 -0.02763428
        8|9
 0.87793812
> cpD <- c(0,pnorm(om1p$zeta - om1p$coef[3]),1); cpD
                     1|2          2|3          3|4          4|5          5|6
0.000000e+00 2.289056e-05 2.251608e-04 2.099568e-03 1.268783e-02 7.157833e-02
         6|7          7|8          8|9
1.799687e-01 4.889769e-01 8.100113e-01 1.000000e+00
> pD<-diff(cpD);pD
         1|2          2|3          3|4          4|5          5|6          6|7
2.289056e-05 2.022702e-04 1.874407e-03 1.058826e-02 5.889050e-02 1.083904e-01
         7|8          8|9
3.090082e-01 3.210344e-01 1.899887e-01
```


additive effect plot
c.taste = 9
c.taste = 8
c.taste = 7
c.taste = 6
c.taste = 5
c.taste = 4
c.taste = 3
c.taste = 2
c.taste = 1
c.taste (probability)
A
B
C
D
additive

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

5-7.3 Housing conditions in Copenhagen

Satisfaction level of housing conditions according to Factor A -Housing –Dwelling type (tower, apartment, atrium, terrace – reference i=1 tower), Factor C –Influence- Feeling of participation in community issues. (low, medium, high - reference j=1 low) and Factor D –Contact- Interaction level with neighboards (low, high - reference k=1 low). N=1681.

5-7.3.1 Nominal response

Deviance table for some models: it is not exhaustive. Best model is the additive one A+ C+D using 14 degrees of freedom and model explicability of 82%.

```
MTB > Name c7 = 'NTRI1' c8 = 'EPROB1' c9 = 'EPROB2' c10 = 'EPROB3' &
CONT>      c11 = 'NOCC1' c12 = 'NOCC2' c13 = 'NOCC3'
MTB > NLogistic 'satisfaction' = housing influence  contact ;
SUBC>   Frequency 'n';
SUBC>   Factors 'housing' 'influence'  'contact';
SUBC>   Reference 'satisfaction' 'low'  &
CONT>     housing 'tower' influence 'low' contact 'low';
SUBC>   Ntrials 'NTRI1';
SUBC>   Eprobability 'EPROB1'-'EPROB3';
SUBC>   Noccur 'NOCC1'-'NOCC3';
SUBC>   Brief 2.

Nominal Logistic Regression: satisfaction versus housing; influence; ...
Response Information
```

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

$$\log \frac{\pi_{ijk2}}{\pi_{ijk1}} = \theta_2 + \alpha_{i2} + \beta_{j2} + \gamma_{k2}$$

$$\alpha_{12} = \beta_{12} = \gamma_{12} = 0$$

$$\log \frac{\pi_{ijk3}}{\pi_{ijk1}} = \theta_3 + \alpha_{i3} + \beta_{j3} + \gamma_{k3}$$

$$\alpha_{13} = \beta_{13} = \gamma_{13} = 0$$

```
Variable  Value              Count
satisfac  low                  567   (Reference Event)
          medium               446
          high                 668
          Total               1681    Frequency:  n
Logistic Regression Table
                                                         Odds        95% CI
Predictor           Coef    SE Coef          Z      P    Ratio     Lower     Upper
Logit 1: (medium/low)
Constant         -0,4192     0,1729    -2,42 0,015
housing
 apartments      -0,4357     0,1725    -2,53 0,012      0,65      0,46      0,91
 atrium           0,1314     0,2231     0,59 0,556      1,14      0,74      1,77
 terraced        -0,6666     0,2063    -3,23 0,001      0,51      0,34      0,77
influenc
 high             0,6649     0,1863     3,57 0,000      1,94      1,35      2,80
 medium           0,4464     0,1416     3,15 0,002      1,56      1,18      2,06
contact
 high             0,3609     0,1324     2,73 0,006      1,43      1,11      1,86
Logit 2: (high/low)
Constant         -0,1387     0,1592    -0,87 0,384
housing
 apartments      -0,7356     0,1553    -4,74 0,000      0,48      0,35      0,65
 atrium          -0,4080     0,2115    -1,93 0,054      0,66      0,44      1,01
 terraced        -1,4123     0,2001    -7,06 0,000      0,24      0,16      0,36
influenc
 high             1,6126     0,1671     9,65 0,000      5,02      3,61      6,96
 medium           0,7349     0,1369     5,37 0,000      2,09      1,59      2,73
contact
 high             0,4818     0,1241     3,88 0,000      1,62      1,27      2,07
Log-likelihood = -1735,042   Test that all slopes are zero: G = 178,794; DF = 12; P-Value = 0,000
Goodness-of-Fit Tests
Method       Chi-Square    DF       P
Pearson          38,910    34   0,258
Deviance         38,662    34   0,267
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

		Deviance Analysis					
	Model	*p*	~~*Deviance or*~~ **Log-likelihood**	*ΔDeviance*	*d.f.*	*Comments* *Contrast*	*Comments* H_0 *Accept.*
0	**1**	2	¿?				
1	**A+C**	12	-1743.072	16.06	2	1 vs 4	No
2	**A+D**	10	-1789.601	109.118	4	2 vs 4	No
3	**C+D**	8	-1766.155	66.226	6	3 vs 4	No
4	**A+C+D**	**14**	**-1735.042**		-	-	-
5	**D+A*C**	26	-1723.764	22.556	12	4 vs 5	No estrict.
6	**C+A*D**	20	-1729.839	10.406	6	4 vs 6	Yes
7	**A+C*D**	18	-1734.447	1.19	4	4 vs 7	Yes

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

5-7.3.2 Ordinal response: logit link using MINITAB

```
MTB > OLogistic 'satisfaction' = housing influence   contact housing *   &
CONT>      influence;
SUBC>   Frequency 'n';
SUBC>   Factors 'housing' 'influence'  'contact';
SUBC>   Logit;
SUBC>   Order 'low' 'medium' 'high';
SUBC>   Reference housing 'tower' influence 'low' contact 'low';
SUBC>   Brief 2.
```

$$\log\frac{\gamma_{ijk1}}{1-\gamma_{ijk1}}=\theta_1+\alpha_i+\beta_j+\alpha\beta_{ij}+\gamma_k$$
$$\alpha_1=\beta_1=\gamma_1=\alpha\beta_{i1}=\alpha\beta_{1j}=0$$

Ordinal Logistic Regression: satisfaction versus housing; influence; ...

```
Link Function:  Logit

Response Information

Variable  Value          Count
satisfac  low              567
          medium           446
          high             668
          Total           1681
Frequency:  n
```

$$\log\frac{\gamma_{ijk2}}{1-\gamma_{ijk2}}=\theta_2+\alpha_i+\beta_j+\alpha\beta_{ij}+\gamma_k$$
$$\alpha_1=\beta_1=\gamma_1=\alpha\beta_{i1}=\alpha\beta_{1j}=0$$

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Const(1)	-0,8882	0,1678	-5,29	0,000			
Const(2)	0,3126	0,1663	1,88	0,060			
housing							
apartments	1,1885	0,1978	6,01	0,000	3,28	2,23	4,84
atrium	0,6067	0,2475	2,45	0,014	1,83	1,13	2,98
terraced	1,6062	0,2415	6,65	0,000	4,98	3,10	8,00
influenc							
high	-0,8689	0,2732	-3,18	0,001	0,42	0,25	0,72
medium	0,1390	0,2124	0,65	0,513	1,15	0,76	1,74
contact							
high	-0,37208	0,09581	-3,88	0,000	0,69	0,57	0,83
housing*influenc							
apartments*high	-0,7198	0,3269	-2,20	0,028	0,49	0,26	0,92

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

```
apartments*medium        -1,0809     0,2654    -4,07 0,000     0,34     0,20     0,57
 atrium*high               0,1556     0,4124     0,38 0,706     1,17     0,52     2,62
 atrium*medium            -0,6511     0,3501    -1,86 0,063     0,52     0,26     1,04
 terraced*high            -0,8446     0,4271    -1,98 0,048     0,43     0,19     0,99
 terraced*medium          -0,8210     0,3319    -2,47 0,013     0,44     0,23     0,84

Log-likelihood = -1728,320
Test that all slopes are zero: G = 192,238; DF = 12; P-Value = 0,000

Goodness-of-Fit Tests

Method      Chi-Square    DF      P
Pearson         25,455    34  0,854
Deviance        25,218    34  0,862

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs           Number  Percent     Summary Measures
Concordant      585603    63,0%     Somers' D                0,31
Discordant      293592    31,6%     Goodman-Kruskal Gamma    0,33
Ties             50371     5,4%     Kendall's Tau-a          0,21
Total           929566   100,0%

MTB > Save  "G:\LIDIA\MLGz2000\MLGZ_01_1\Exemples_teo\RP_ex6_habi.mpj";
SUBC>    Project;
SUBC>    Replace.
Saving file as: G:\LIDIA\MLGz2000\MLGZ_01_1\Exemples_teo\RP_ex6_habi.mpj
* NOTE  * Existing file replaced.
MTB > OLogistic 'satisfaction' = housing influence   contact ;
SUBC>   Frequency 'n';
SUBC>   Factors 'housing' 'influence'  'contact';
SUBC>   Logit;
SUBC>   Order 'low' 'medium' 'high';
SUBC>   Reference housing 'tower' influence 'low' contact 'low';
SUBC>   Brief 2.
```

Ordinal Logistic Regression: satisfaction versus housing; influence; ...

```
Link Function:  Logit
Response Information
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

```
Variable  Value             Count
satisfac  low                 567
          medium              446
          high                668
          Total              1681
Frequency:  n
Logistic Regression Table
                                                       Odds       95% CI
Predictor         Coef    SE Coef        Z      P     Ratio    Lower    Upper
Const(1)       -0,4961     0,1245    -3,98 0,000
Const(2)        0,6907     0,1252     5,52 0,000
housing
 apartments     0,5723     0,1187     4,82 0,000      1,77     1,40     2,24
 atrium         0,3662     0,1568     2,34 0,019      1,44     1,06     1,96
 terraced       1,0910     0,1515     7,20 0,000      2,98     2,21     4,01
influenc
 high          -1,2888     0,1267   -10,17 0,000      0,28     0,21     0,35
 medium        -0,5664     0,1050    -5,40 0,000      0,57     0,46     0,70
contact
 high         -0,36028    0,09536    -3,78 0,000      0,70     0,58     0,84

Log-likelihood = -1739,575
Test that all slopes are zero: G = 169,728; DF = 6; P-Value = 0,000

Goodness-of-Fit Tests
Method      Chi-Square    DF      P
Pearson         47,887    40  0,183
Deviance        47,728    40  0,187
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

5-7.3.3 Ordinal response: logit link using R (final model only)

```
> summary(copen)
       id              housing     influence   contact    satisfaction        n
 Min.   : 1.00   tower      :18   low    :24   low :36   low    :24    Min.   : 3.00
 1st Qu.:18.75   apartments:18   medium:24   high:36   medium:24    1st Qu.:10.00
 Median :36.50   atrium     :18   high   :24             high   :24    Median :19.50
 Mean   :36.50   terraced   :18                                        Mean   :23.35
 3rd Qu.:54.25                                                         3rd Qu.:31.75
 Max.   :72.00                                                         Max.   :86.00
> library(MASS)
> copen.polr <- polr(satisfaction~housing*influence+contact,data=copen,weights=n)
> summary(copen.polr)

Call: polr(formula = satisfaction ~ housing * influence + contact, data = copen, weights = n)

Coefficients:
                                    Value Std. Error t value
housingapartments                 -1.1885    0.19724 -6.0256
housingatrium                     -0.6067    0.24457 -2.4808
housingterraced                   -1.6062    0.24100 -6.6650
influencemedium                   -0.1390    0.21255 -0.6541
influencehigh                      0.8689    0.27434  3.1671
contacthigh                        0.3721    0.09599  3.8764
housingapartments:influencemedium  1.0809    0.26585  4.0657
housingatrium:influencemedium      0.6511    0.34500  1.8873
housingterraced:influencemedium    0.8210    0.33067  2.4829
housingapartments:influencehigh    0.7198    0.32873  2.1896
housingatrium:influencehigh       -0.1555    0.41048 -0.3789
housingterraced:influencehigh      0.8446    0.43027  1.9630

Intercepts:
            Value   Std. Error t value
low|medium  -0.8882  0.1672    -5.3135
medium|high  0.3126  0.1657     1.8871

Residual Deviance: 3456.64
AIC: 3484.64
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

Non exhaustive deviance table. Best model is D +A*C , it includes interaction between housing type and influence, using 14 degrees of freedom and model explicability is 78%, residual deviance is 25.22 with 48-14=34 d.f, leading to a goodness of fit p-value of 0.86 (Null hypothesis 'model is consistent to data').

		Deviance Analysis					
	Model	*p*	***Log-likelihood***	*ΔDeviance*	*d.f.*	*Comments*	
						Contrast	H_0 *Accept.*
0	**1**	2	¿?				
1	**A+C**	7	-1746.728	14.306	1	1 vs 4	No
2	**A+D**	6	-1793.694	108.238	2	2 vs 4	No
3	**C+D**	5	-1767.53	55.91	3	3 vs 4	No
4	**A+C+D**	8	-1739.575	-	-	-	-
5	**D+A*C**	**14**	**-1728.320**	**22.51**	**6**	**4 vs 5**	**No**
6	**C+A*D**	11	-1735.242	8.666	3	4 vs 6	Yes
7	**A+C*D**	10	-1739.470	0.21	2	4 vs 7	Yes

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

Proportional odds model gives cut-points for the latent variable formulation of -0.89 and 0.31, that correspond to cumulative odds of 0.41 and 1.37 and cumulative probabilities of 0.29 and 0.58 for reference group (block of appartments, influence low and contact low): 29% indicating low satisfaction, 29% with medium satisfaction (58-29) and 42% (100-58) having high satisfaction.

A negative coefficient indicates a shift along lower latent variable scales (latent variable formulation). It means increasing cumulative logodds and thus it models an increment in lower satisfaction probability and a decrement in high satisfaction probability. Keep in mind that latent variable signs and generalized linear model signs applied to cumulative probabilities are opposed.

$$j=1 \quad \log\frac{\gamma_{i1}}{1-\gamma_{i1}} = \log\frac{\pi_{i1}}{1-\pi_{i1}} = \alpha_1 + \mathbf{x_i}^{\mathbf{T}}\beta = -\log\frac{1-\pi_{i1}}{\pi_{i1}} \rightarrow \log\frac{1-\pi_{i1}}{\pi_{i1}} = \log\frac{\pi_{i2}+\pi_{i3}}{\pi_{i1}} = \log\frac{1-\gamma_{i1}}{\gamma_{i1}} = -\alpha_1 - \mathbf{x_i}^{\mathbf{T}}\beta$$

$$j=2 \quad \log\frac{\gamma_{i2}}{1-\gamma_{i2}} = \log\frac{\pi_{i1}+\pi_{i2}}{\pi_{i3}} = \alpha_1 + \mathbf{x_i}^{\mathbf{T}}\beta = -\log\frac{\pi_{i3}}{1-\pi_{i3}} \rightarrow \log\frac{\pi_{i3}}{1-\pi_{i3}} = -\alpha_2 - \mathbf{x_i}^{\mathbf{T}}\beta$$

Factor D (Contact) parameter estimate is positive (latent scale), indicating that residents having social interaction with their neighboars are generally more satisfied than isolated ones. Satisfection odds are 45% greater for high Factor D than low Factor D. The same applies to 'medium' logodds or 'high' logodds over 'low' satisfaction. This is the trend shown by the model.

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

To interpret Factor A and C effects, interactions have to considered. For example, residents showning a low level Influence (Factor C) needs to include Factor A (Dwelling type) characteristics, apartment, atrium and terrace dwelling negatively affect to satisfaction level (target) compared to 'block' dwelling. In general, participation in the comunity (Factor C) increments housing satisfaction. Having a high Influence (Factor) instead of 'low' means a 57% decrement in satisfaction odds of 'low' or 'low plus medium' over 'high' for block unit residents, 80% for apartment residents, 51% for 'atrium houses' and 82% for 'terraced houses'. Having a 'medium' Influence level (Factor C) is generally better than 'low' (with the exception of block residents), but not as much good as having an influence level 'high' (with exception of 'atrium houses').

5-7.3.4 Ordinal response: probit link

Ordinal probit model relies on transforming cumulative satisfaction probabilities using probit link and considering a linear model on explanatory variables. A latent variate formulation is the most natural option, assuming standard normal distribution for the latent variable. Ordinal logit and probit estimates for model parameters are similar once reescaled logit estimates by dividing into standard logit estándar deviation. Ordinal probit model assumes a latent variate scale equal to 1, while the equivalent ordinal logit model assumes a standard deviation of the latent variate scale of $\pi/\sqrt{3}$. Model estimates according to MINITAB are shown below and can be interpreted, once signs are changed, as an ordinary linear regression in the linear predictor scale.

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

Let us consider D+A*C model only.

```
MTB > OLogistic 'satisfaction' = housing influence   contact housing* &
CONT>      influence;
SUBC>   Frequency 'n';
SUBC>   Factors 'housing' 'influence'  'contact';
SUBC>   Normit;
SUBC>   Order 'low' 'medium' 'high';
SUBC>   Reference housing 'tower' influence 'low' contact 'low';
SUBC>   Brief 2.
```

Ordinal Logistic Regression: satisfaction versus housing; influence; ...

```
Link Function:  Normit

Response Information

Variable  Value        Count
satisfac  low            567
          medium         446
          high           668
          Total         1681
Frequency:  n

Logistic Regression Table

Predictor                 Coef    SE Coef        Z      P
Const(1)               -0,5440     0,1025    -5,31 0,000
Const(2)                0,1892     0,1020     1,85 0,064
housing
 apartments             0,7281     0,1205     6,04 0,000
 atrium                 0,3721     0,1519     2,45 0,014
 terraced               0,9790     0,1458     6,71 0,000
influenc
 high                  -0,5165     0,1633    -3,16 0,002
 medium                 0,0864     0,1302     0,66 0,507
contact
 high                 -0,22846    0,05832    -3,92 0,000
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

```
housing*influenc
 apartments*high       -0,4479     0,1963    -2,28 0,023
 apartments*medium     -0,6600     0,1624    -4,06 0,000
 atrium*high            0,0780     0,2500     0,31 0,755
 atrium*medium         -0,4108     0,2149    -1,91 0,056
 terraced*high         -0,5217     0,2578    -2,02 0,043
 terraced*medium       -0,4964     0,2022    -2,45 0,014

Log-likelihood = -1728,665
Test that all slopes are zero: G = 191,547; DF = 12; P-Value = 0,000

Goodness-of-Fit Tests
Method      Chi-Square    DF      P
Pearson         26,313    34  0,824
Deviance        25,909    34  0,839

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs           Number  Percent     Summary Measures
Concordant      586041    63,0%     Somers' D                0,31
Discordant      294028    31,6%     Goodman-Kruskal Gamma    0,33
Ties             49497     5,3%     Kendall's Tau-a          0,21
Total           929566   100,0%

MTB >
```

Residual model deviance is 25.9 on 34 d.f., very similar to the value obtained by the ordinal logit model.

Cut points have to be interpreted according to z, standard normal distribution values: the limit between 'low' and 'medium' satisfaction lies on z=-0.54 and the limit 'medium' and 'high' lies on z=0.19. These values give cumulative probabilities of ($\Phi(-0.54)=0.29$) and $\Phi(0.19)=0.58$: 29% show a 'low' satisfaction, 29% a 'medium' satisfaction level and 42% have ´high´ satisfaction.

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

To interpret Factor D (Contact) estimates, residents with high relation with their neighbours show a latent scale satisfaction of 0.23, being this estimate very similar to the one obtained by the ordinal logit model. Ordinal logit latent scale estimate was 0.37 with standard deviation (standard logit) $\pi/\sqrt{3}=1.81$, dividing estimate by logit standard deviation returns 0.37/1.81=0.21 precisice ordinal probit estimate in this last model.

5-7.3.5 Ordinal response: link clog_log

This third option is based on a log-log complementary link and gives the model formulation:

$$\mathbf{log}\left(-\mathbf{log}\left(1-\gamma_{ij}\right)\right)=\boldsymbol{\theta}_j+\mathbf{x}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta}$$

This model can be interpreted in the latent variate approach with an extreme type distribution (log-Weibull or Gompertz/Gumbel) with location parameter 0 and scale parameter 1 and distribution function: $F(\eta)=1-\mathbf{exp}(-\mathbf{exp}(\eta))$.

This is an assymetric distribution with expected value equal to Euler constant –0.57722 and variance $\pi^2/6$. Median is loglog2.

Let us consider D+A*C model.

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

```
MTB > OLogistic 'satisfaction' = housing influence   contact housing* &
CONT>      influence;
SUBC>   Frequency 'n';
SUBC>   Factors 'housing' 'influence'  'contact';
SUBC>   Gompit;
SUBC>   Order 'low' 'medium' 'high';
SUBC>   Reference housing 'tower' influence 'low' contact 'low';
SUBC>   Brief 2.
```

Ordinal Logistic Regression: satisfaction versus housing; influence; ...

```
Link Function:  Gompit

Response Information
Variable  Value        Count
satisfac  low            567
          medium         446
          high           668
          Total         1681
Frequency:  n

Logistic Regression Table

Predictor                 Coef    SE Coef        Z      P
Const(1)               -1,0332     0,1245    -8,30  0,000
Const(2)               -0,1760     0,1210    -1,46  0,146
housing
 apartments             0,7669     0,1369     5,60  0,000
 atrium                 0,4264     0,1729     2,47  0,014
 terraced               1,0052     0,1556     6,46  0,000
influenc
 high                  -0,6711     0,2201    -3,05  0,002
 medium                 0,0793     0,1555     0,51  0,610
contact
 high                 -0,21452    0,06558    -3,27  0,001
```

POLYTOMOUS TARGETS. EXAMPLE MADSEN-76, AGRESTI-90

```
housing*influenc
 apartments*high       -0,4191     0,2555    -1,64 0,101
 apartments*medium     -0,6916     0,1873    -3,69 0,000
 atrium*high            0,1651     0,3142     0,53 0,599
 atrium*medium         -0,4165     0,2496    -1,67 0,095
 terraced*high         -0,4416     0,3199    -1,38 0,167
 terraced*medium       -0,4652     0,2181    -2,13 0,033

Log-likelihood = -1732,930
Test that all slopes are zero: G = 183,017; DF = 12; P-Value = 0,000

Goodness-of-Fit Tests

Method      Chi-Square    DF      P
Pearson         35,766    34  0,385
Deviance        34,439    34  0,447
```

Residual deviance is 34.439 with 34 d.f., not as good as the value obtained with the previous proposals, ordinal logit and probit models.

Cut-points are –1.0332 and –0.176, that correspon to cumulative probabilities for the reference group of 0.299 and 0.568: nearly 30% of the reference group shows a 'low' satisfaction, almost 27% a 'medium' satisfaction and 43% a 'high' satisfaction.

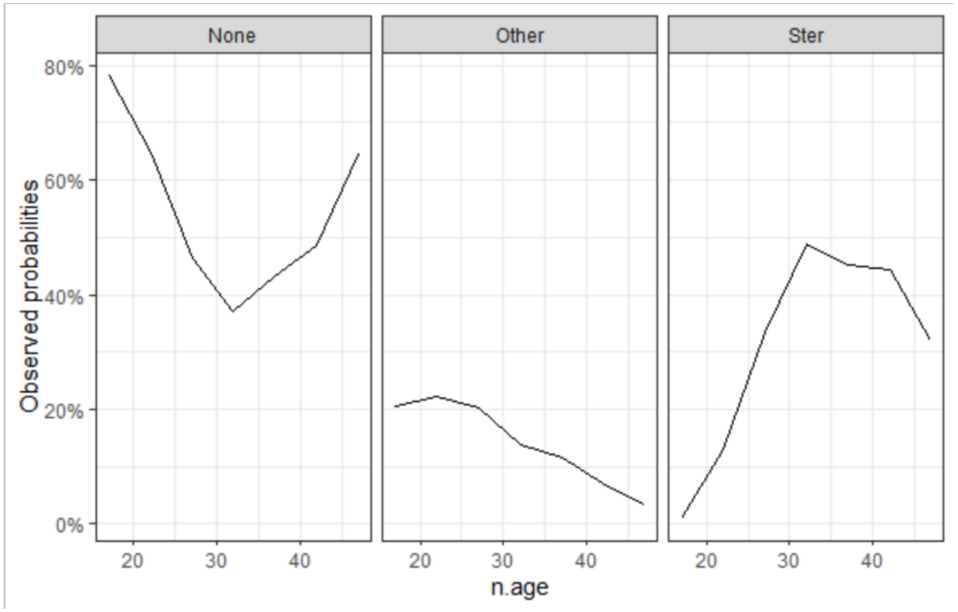
POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

5-7.4 Fesal (1985)

Elaborated by German Rodríguez from the final report of the Demographic and Health Survey conducted in El Salvador in 1985 (FESAL-1985). Data table shows 3165 currently married women classified by age, grouped in five-year intervals, and current use of contraception, classified as sterilization, other methods, and no method.

Current Use of Contraception By Age Currently Married Women. El Salvador, 1985

c.age	Contraceptive Method			All	n.age
	Ster.	Other	None		
15-19	3	61	232	**296**	**17**
20-24	80	137	400	**617**	**22**
25-29	216	131	301	**648**	**27**
30-34	268	76	203	**547**	**32**
35-39	197	50	188	**435**	**37**
40-44	150	24	164	**338**	**42**
45-49	91	10	183	**284**	**47**
All	**1005**	**489**	**1671**	**3165**	



POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

5-7.4.1 Nominal Outcome modelling

```
> summary(mm4m)
Call: multinom(formula = c.use ~ n.age + I(n.age^2) + I(n.age^3), data = fesal85, weights = y)
Coefficients:
      (Intercept)     n.age    I(n.age^2)    I(n.age^3)
Other   -3.245387 0.1250483  0.0001217356 -5.611951e-05
Ster   -23.362462 1.7651664 -0.0424949026  3.237379e-04
Std. Errors:
       (Intercept)        n.age   I(n.age^2)   I(n.age^3)
Other 8.342487e-07 1.667839e-05 0.0002626260 7.207507e-06
Ster  4.792138e-07 1.100847e-05 0.0001959075 4.853262e-06
Residual Deviance: 5750.261
AIC: 5766.261
> anova( mm3, mm4, test="Chisq")
Likelihood ratio tests of Multinomial Models
Response: c.use
           Model Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1 poly(n.age, 2)        36   5766.273
2 poly(n.age, 3)        34   5750.260 1 vs 2     2 16.01257 0.0003333602
> anova( mm2, mm3, test="Chisq")
Likelihood ratio tests of Multinomial Models
Response: c.use
           Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
1          n.age        38   6039.776
2 poly(n.age, 2)        36   5766.273 1 vs 2     2 273.5031       0
```

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

```
> summary(mm1)
Call:
multinom(formula = c.use ~ g.age, data = fesal85, weights = y)
…

> AIC(mm1,mm4,mm2,mm3,mm4m)
     df      AIC
mm1  14 5773.798   # Y ~ g.age
mm4   8 5766.260   # Y ~ poly(n.age,3)
mm2   4 6047.776   # Y ~ n.age
mm3   6 5778.273   # Y ~ poly(n.age,2)
mm4m  8 5766.261   # Y ~ n.age + I(n.age^2) + I(n.age^3)
```

$$\log \frac{\pi_{iOther}}{\pi_{iNone}} = -3.245387 + 0.1250483\ \text{n.age} + 0.0001217356\ \text{n.age}^2 - 5.611951\text{e-}05\ \text{n.age}^3$$

$$\log \frac{\pi_{iSter}}{\pi_{iNone}} = -23.362462 + 1.7651664\ \text{n.age} - 0.0424949026\ \text{n.age}^2 + 3.237379\text{e-}04\ \text{n.age}^3$$

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

5-7.4.2 Ordinal Outcome modelling

```
> om1 <- polr( c.use ~ g.age, weight=y, data=fesal85)
> om2 <- polr( c.use ~ n.age, weight=y, data=fesal85)
> om3 <- polr( c.use ~ poly(n.age,2), weight=y, data=fesal85)
> om4 <- polr( c.use ~ poly(n.age,3), weight=y, data=fesal85)
> om3m <- polr( c.use ~ n.age + I((n.age)^2), weight=y, data=fesal85)
> anova( om2, om3, test="Chisq" )
Likelihood ratio tests of ordinal regression models
Response: c.use
           Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
1          n.age      3162   6167.992
2 poly(n.age, 2)      3161   5972.486 1 vs 2     1 195.5066       0
> anova( om3, om4, test="Chisq" )
Likelihood ratio tests of ordinal regression models
Response: c.use
           Model Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
1 poly(n.age, 2)      3161   5972.486
2 poly(n.age, 3)      3160   5971.436 1 vs 2     1 1.050122 0.3054791
> AIC( mm4, om3m, om1, om3mp, om3mg )
      df      AIC
mm4    8 5766.260
om3m   4 5980.486 # logit n.age + I((n.age)^2)
om1    8 5979.335 # g.age
om3mp  4 5961.966 # probit link
om3mg  4 5914.791 # gompit link
```

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

```
> summary(om3m)
Call:
polr(formula = c.use ~ n.age + I((n.age)^2), data = fesal85, weights = y)

Coefficients:
                 Value Std. Error t value
n.age         0.461065  0.0063369   72.76
I((n.age)^2) -0.006565  0.0001884  -34.84

Intercepts:
            Value      Std. Error t value
None|Other     7.5665     0.0006 12852.3169
Other|Ster     8.2720     0.0294   281.1563

Residual Deviance: 5972.486
AIC: 5980.486
```

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

```
> summary(om3mp)
Call: polr(formula = c.use ~ n.age + I((n.age)^2), data = fesal85, weights = y, method =
"probit")
Coefficients:
                  Value Std. Error t value
n.age          0.286667  3.168e-03   90.48
I((n.age)^2) -0.004061  8.875e-05  -45.76

Intercepts:
            Value      Std. Error t value
None|Other     4.7308     0.0004 11447.4171
Other|Ster     5.1666     0.0179   288.5634

Residual Deviance: 5953.966
AIC: 5961.966
```

$$log\left(\frac{\gamma_{iNone}}{1-\gamma_{iNone}}\right) = 7.567 - 0.461n.age + 0.0066\,n.age^2$$

$$log\left(\frac{\gamma_{iOther}}{1-\gamma_{iOther}}\right) = 8.272 - 0.461n.age + 0.0066\,n.age^2$$

or

$$probit(\gamma_{ij}) = \begin{Bmatrix} 4.7308 \\ 5.1666 \end{Bmatrix} - 0.286667n.age + 0.004061\,n.age^2$$

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

5-7.4.3 Hierarchical modelling

```
> fesal85$bc.use <- factor(ifelse(fesal85$c.use=="None",0,1), labels = c("None", "Some" ))
> bh1m0 <- glm( bc.use ~ 1, family = binomial, data = fesal85, weights = y )
> bh1m1 <- glm( bc.use ~ n.age, family = binomial, data = fesal85, weights = y )
> bh1m2 <- glm( bc.use ~ poly(n.age,2), family = binomial, data = fesal85, weights = y )
> bh1m3 <- glm( bc.use ~ poly(n.age,3), family = binomial, data = fesal85, weights = y )
> anova( bh1m1, bh1m2, test="Chisq" )
Analysis of Deviance Table

Model 1: bc.use ~ n.age
Model 2: bc.use ~ poly(n.age, 2)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1        19     4339.7
2        18     4165.0  1   174.74 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova( bh1m2, bh1m3, test="Chisq" )
Analysis of Deviance Table

Model 1: bc.use ~ poly(n.age, 2)
Model 2: bc.use ~ poly(n.age, 3)
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        18     4165.0
2        17     4164.9  1 0.077617   0.7806
```

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

POLYTOMOUS TARGETS. EXAMPLE FESAL 1985

```
> # Prediction 22 years
> bh1 <- glm( bc.use ~ I((n.age-22)) + I((n.age-22)^2), family = binomial, data = fesal85,
weights = y )
> bh2 <- glm( c.use ~ I((n.age-22)) + I((n.age-22)^2) + I((n.age-22)^3), family = binomial,
data = fesal85[fesal85$bc.use=="Some",], weights = y )
> coef(bh1)
      (Intercept)   I((n.age - 22)) I((n.age - 22)^2)
     -0.498244515       0.154226121      -0.006344751
> coef(bh2)
      (Intercept)   I((n.age - 22)) I((n.age - 22)^2) I((n.age - 22)^3)
    -0.6182446111      0.3172773783     -0.0176353060      0.0003828289
```

$$log\left(\frac{\pi_{iSome}}{1-\pi_{iSome}}\right) = -0.498 + 0.154(n.age-22) - 0.0063\,(n.age-22)^2$$

$$log\left(\frac{\pi_{iSter}}{1-\pi_{iSter}}\right) = -0.618 + 0.317(n.age-22) - 0.0176(n.age-22)^2 + 0.00038(n.age-22)^3$$