

Ordinal regression models made easy. A tutorial on parameter interpretation, data simulation, and power analysis.

Filippo Gambarota¹ & Gianmarco Altoè¹

¹ Department of Developmental Psychology and Socialization, University of Padova, Italy

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. Filippo Gambarota: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Gianmarco Altoè: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Filippo Gambarota, Postal address. E-mail: filippo.gambarota@unipd.it

Abstract

Abstract here

Keywords: ordinal, likert, simulations, power

Word count: X

Ordinal regression models made easy. A tutorial on parameter interpretation, data simulation, and power analysis.

Introduction

Psychological research make an extensive use of ordinal data. One of the main reason is probably the usage of Likert scales (Likert, 1932). Ordinal data refers to a specific type of measurement scale (Stevens, 1946) where ordered numbers are assigned to a variable. Compared to nominal scale and as the name suggest the labels are ordered. Compared to interval or ratio scales there is no explicit assumption about the distance between labels. (rivedi questo). An example is asking people the degree of agreement about a certain statement using a scale from 1 (no agreement) to 7 (total agreement). Answering 4 compared to 2 suggest an higher agreement but we cannot affirm that there is two times the agreement compared to the second answer. Stevens (1946) and Kemp and Grace (2021) suggested that for ordinal variables is appropriate to calculate ranks-based statistics (e.g., median or percentiles) instead of metric statistics (e.g., mean or standard deviation). This distinction in terms of the appropriateness of certain descriptive statistics is also relevant when modeling data. Treating ordinal data as metric refers to assuming the labels as actual integer numbers thus assuming a fixed and know distance between levels (Liddell & Kruschke, 2018).

In Psychology especially when using item-based measures (questionnaires, surveys, etc.) the common practice is using a normal linear regression that makes an explicit assumption about metric features of the response variable. Liddell and Kruschke (2018) reviewed the psychological literature using likert-based measures and reported how the majority of papers used metric-based statistical models. In the same work, Liddell and Kruschke (2018) showed extensive examples and simulations about the potential pitfalls of treating an ordinal variable as metric (but see Robitzsch, 2020 for an alternative perspective). They reported problems in terms of lack of power, inversion of the effects and

distortion in estimating the effect size (rivedi meglio qui). (vedi se c'è qualche altro lavoro che fa vedere questa cosa). Some authors suggested that despite individual items being ordinal, averaging multiple items could solve the problem of applying metric models (e.g., Carifio & Perla, 2007, 2008 metti queste citazioni). Liddell and Kruschke (2018) suggested that in some conditions even averaging multiple items and applying metric models could be problematic. Cliff (2016) suggest that most of the research questions in behavioral sciences can be considered as ordinal (*is the score x higher than the score y ?*) concerning variables where the most appropriate measurement scale is probably ordinal.

qualcosa ancora qui

Ordinal regression models

Despite the actual modeling proposal by Liddell and Kruschke (2018), there is a class of regression models taking into account the ordinal nature of the response variable without making metric assumptions. We can class this general class of models as *ordinal regression*. The nomenclature of these models can be confusing mainly because there are several types of models with different assumptions and structures (Gerhard Tutz, 2022). Gerhard Tutz (2022) and Bürkner and Vuorre (2019) provide a clear and updated taxonomy of ordinal regression models. We can identify three main classes: *cumulative models* (McCullagh, 1980), *sequential models* (Gerhard Tutz, 1990) and *adjacent category models*. The cumulative is the mostly used model assuming the existence of a latent variable that when segmented using thresholds produces the observed ordinal variable. The psychological process underlying the response is clearly formalized in the signal detection theory framework where the respondent (vedi come sistemare). The sequential model as suggested by the name is appropriate when modelling sequential processes. Assuming to have 5 response options, the sequential model assume that responding 3 means already reaching the states 1 and 2. A clear example is proposed by Bürkner and Vuorre (2019) where the marriage duration in years is predicted as a function of some explanatory variables. For

each level of the response variable there is a latent distribution where the step between a marriage year $k = 1$ and the next years $k > 1$ is modeled by the sequential regression. When comparing k with $k > 1$, everything lower than k is assumed to be already reached (Gerhard Tutz & Berger, 2020). The adjacent category model compare the category k with $k + 1$ still assuming a latent distribution for each k . As suggested by Gerhard Tutz (2022) the adjacent-category model can be seen as a series of binary binomial regressions taking into account the order of the categories. Bürkner and Vuorre (2019) suggested that adjacent-category model can be chosen for its mathematical convenience and there is no a clear empirical distinction as for the cumulative vs sequential model.

In the current paper we put the focus on the cumulative model for several reasons. The first reason is that the latent formulation of the model is particularly convenient both for parameter interpretation and data simulation. The second reason is that several psychological variables can be formalized as a latent continuous construct that when measured is collected using ordinal items. Furthermore, also the signal detection theory framework models the decision process assuming a latent distribution for the signal and noise where thresholds are the response criteria (e.g., DeCarlo, 2010).

Model notation

metti qualcosa sulla nomenclatura

The name, cumulative link models is adopted from Agresti (2002), but the model class has been referred to by several other names in the literature, such as ordered logit models and ordered probit models (Greene and Hensher 2010) for the logit and probit link functions. The cumulative link model with a logit link is widely known as the proportional odds model due to McCullagh (1980) and with a complementary log-log link, the model is sometimes referred to as the proportional hazards model for grouped survival times.

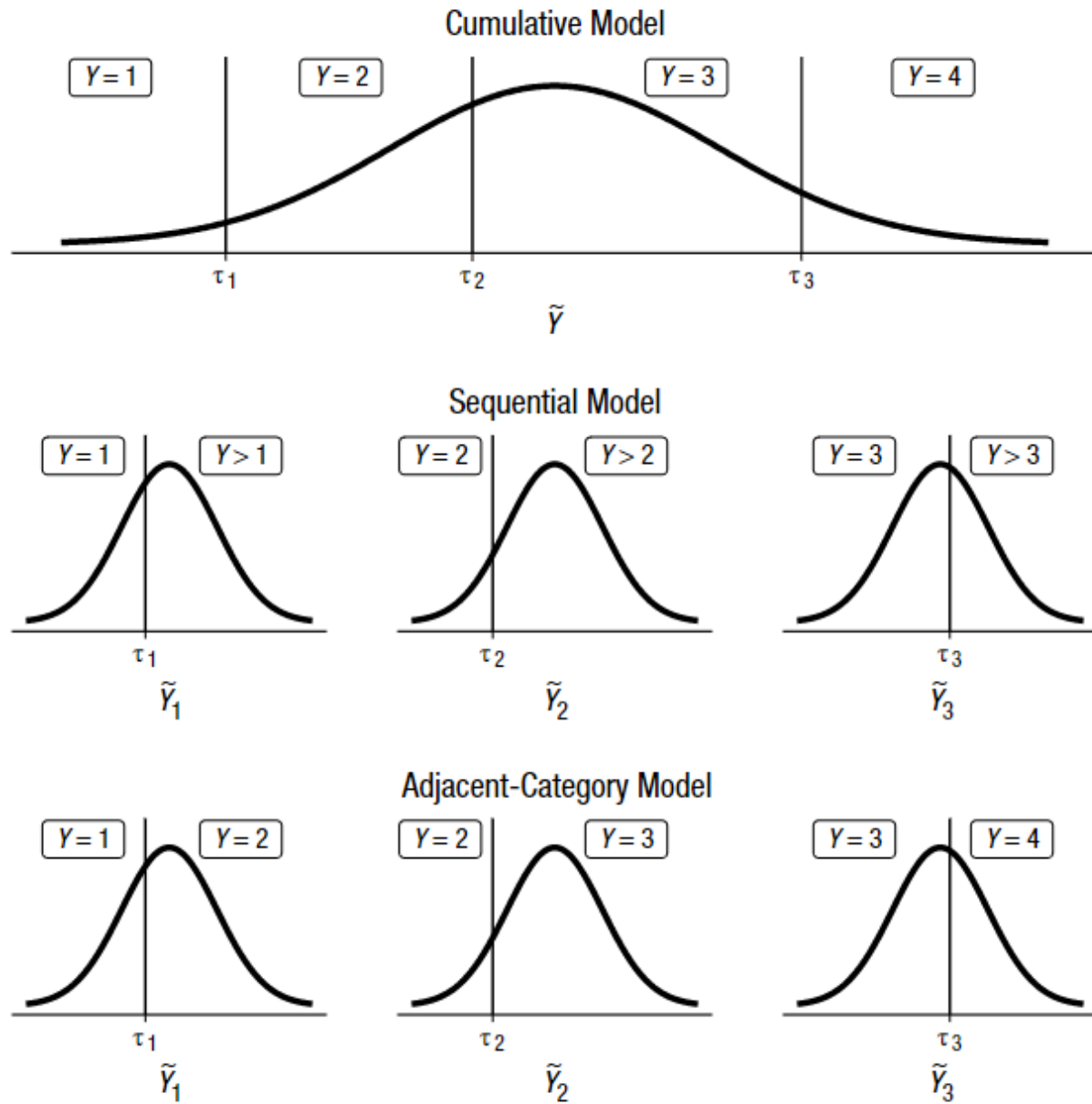


Figure 1. Theoretical figure about ordinal models. Voglio fare una figura come questa in Bürkner and Vuorre (2019) ma con qualche modifica per fare vedere i tre modelli.

In this section we introduce some notations for the cumulative model that is used through the paper and in the R code. We proposed a notation as consistent as possible with the main textbooks and seminal papers and with the `ordinal` R package used in the tutorial. We define Y_k as the observed ordinal response with $1, \dots, k$ levels and Y^* is the underlying latent variable. The latent variable is segmented using $k - 1$ thresholds $\alpha_1, \dots, \alpha_{k-1}$. Similarly to the generalized linear models framework, we define $g(x) = \eta$ as the link function that maps that transforms probabilities into the linear predictor η . To transform η into probabilities we use the inverse of the link function $x = g^{-1}(\eta)$. The specific link function define the type of model and require a different R function. For example, when assuming a Gaussian distribution we are fitting a model with a `probit` link function is the cumulative distribution function $g(x) = \Phi^{-1}(x) = \eta$ and the inverse of the link function is the inverse cumulative distribution function (or quantile function) defined $x = g^{-1}(\eta) = \Phi(\eta)$. When modelling an ordinal variable in a cumulative link model we actually modelling the cumulative probability $P(Y \leq k), k = 1, \dots, k - 1$. Equation (1) depict the general cumulative model including predictors \mathbf{X} and regression coefficients β . The minus sign in $\mathbf{X}\beta$ is used to give the usual interpretation of standard regression models where an increase in β corresponds to an increase in the probability of higher categories of Y (metti il libro di agresti sui dati ordinali).

$$P(Y \leq k) = g^{-1}(\alpha_k - \mathbf{X}\beta), \quad k = 1, \dots, k - 1 \quad (1)$$

The $\mathbf{X}\beta$ is the linear predictor η that is the cumulative probability $P(Y \leq k)$ transformed using the link function $g()$. To obtain the probability of a single outcome $P(Y = k)$ we can compute the difference between cumulative probabilities as shown in Equation (??).

$$P(Y = k) = g^{-1}(\alpha_k - \eta) - g^{-1}(\alpha_{k-1} - \eta), \quad k = 1, \dots, k - 1 \quad (2)$$

There are always two special cases when computing the probability of a single outcome Y that is when $Y = 1$ and $Y = k$. In the first case the cumulative probability is calculated from $-\infty$ that is the same as temporary assuming an $\alpha_0 = -\infty$. In the second case ($Y = 1$) the probability is calculated as $P(Y = k) = 1 - g^{-1}(\alpha_{k-1} - \eta)$ that is the same as assuming a temporary threshold $\alpha_k = +\infty$. The Figure 2 shows how the single probabilities of the ordinal outcome are calculated from cumulative probabilities.

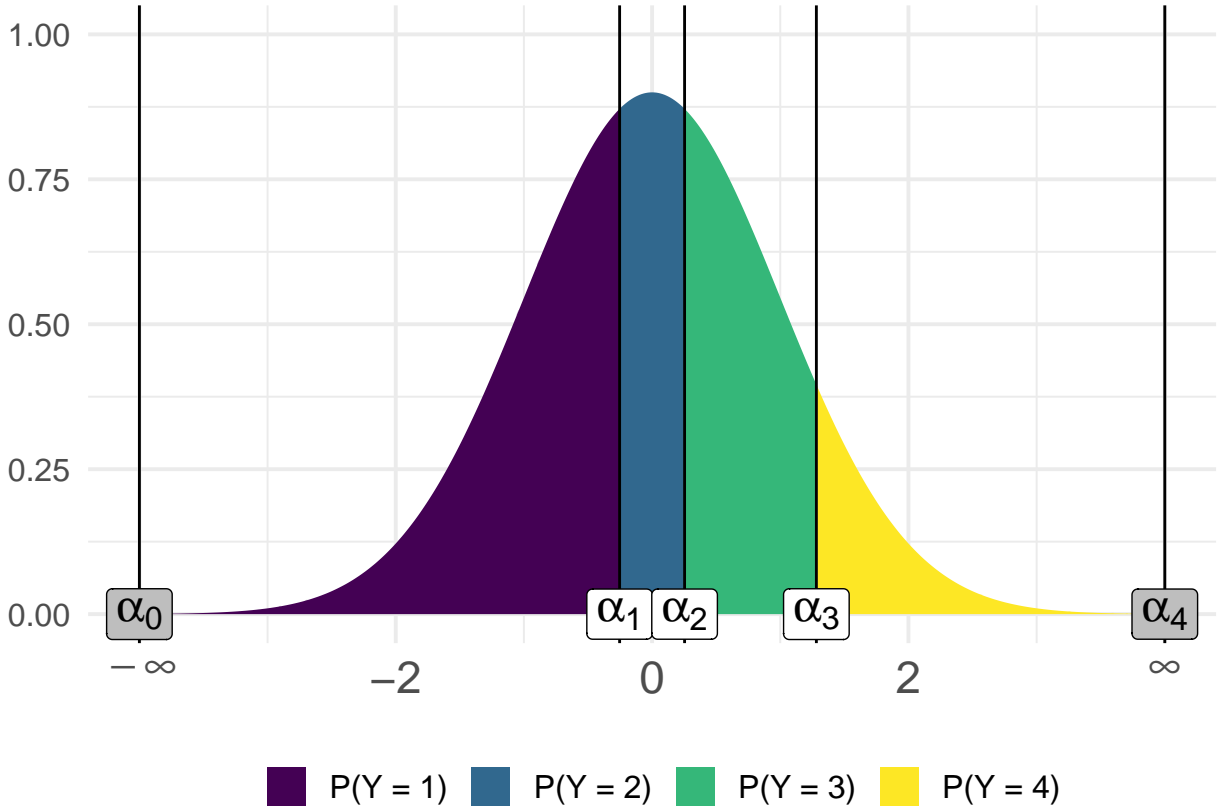


Figure 2. caption here

The model presented in Equation (1) can be considered similar to a standard binomial generalized linear model. In this case we have $k - 1$ equations, one for each threshold α and we model the cumulative probability. The same model can be written as in Equation (3) assuming a continuous latent variable Y^* explained by the linear combination of predictors $\eta = \mathbf{X}\beta$ as in a standard linear regression model.

$$Y_i^* = \eta + \epsilon_i \quad (3)$$

The crucial part is ϵ_i that is the random part of the model coming from a certain probability distribution. For a *probit* model, errors are sampled from a standard Gaussian distribution while for a *logit* model from a standard logistic distribution. Following the notation by Gerhard Tutz (2022), the observed ordinal value $Y_i = k$ comes from Y_i^* belonging to the interval defined by the thresholds $Y_i = k \iff \alpha_{k-1} < Y_i^* < \alpha_k$ where $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_{k-1} < \alpha_k = \infty$.

In the basic version of the model, the thresholds α_k are considered as fixed and being part of the measurement procedure (Liddell & Kruschke, 2018) and do not vary as a function of the predictors. In a more sophisticated version of the model called location-shift (Gerhard Tutz, 2022), both the location μ and the thresholds α_k can vary as a function of the predictors. Gerhard Tutz (2022) described also another version of the model called location-scale (Cox, 1995; Rigby & Stasinopoulos, 2005; Gerhard Tutz, 2022) where the location μ and the scale σ^2 of the distribution can vary as a function of the predictors.

Alternative parametrizations

Several parametrizations of the ordinal models have been proposed by different authors. Liddell and Kruschke (2018) and Kruschke (2015) proposed a Bayesian *probit* model where the latent mean and standard deviations are estimated along with the threshold. The interesting part is that the thresholds are transformed to be on the scale of the ordinal variable (e.g, from 1 to 4 when $k = 4$) and not in terms of z scores. Kurz (2023) explain how to convert a model fitted with the **brms** package (citation) (an R package for regression modeling using *stan*) into the Kruschke (2015) parametrization. Gelman, Hill, and Vehtari (2020) proposed three alternative parametrizations focusing on different definition of the thresholds. The current tutorial is focused on the standard

parametrization of the cumulative link model as described in the equations above.

Link function

The cumulative link model implemented in Equations (1) and (3) can be considered the general formulation that requires specifying the link function $g(\cdot)$ or the latent distribution $\epsilon_i \sim D(\mu, \sigma^2)$. Among several available functions the *logit* and *probit* models are the most common. The *logit* model assume a *logit* link function and a logistic distribution as latent variable. On the other side, the *probit* model assume a Gaussian distribution. The two models provide similar results but the parameters interpretation and the way data can be simulated is different. In the next sections we will illustrate the differences and simulation strategies. Figure 3 depicts the two distributions while Table 1 summarise the presented cumulative models with the proposed link function and the corresponding R code. In terms of parameters, both distributions can be defined with a location μ and a scale s parameter. The standard normal distribution has $\mu = 0$ and $s = 1$. Furthermore the variance corresponds to the scale $s^2 = \sigma^2 = 1$. The variance of the logistic distribution is $\sigma^2 = \frac{s^2\pi^2}{3}$. The standard logistic distribution has $\mu = 0$ and $s^2 = 1$ thus the standard deviation simplified to $\frac{\pi}{\sqrt{3}} \approx 1.81$. In practical terms, the standard version of the two distributions are very similar with the logistic having more variance.

Table 1

caption here

Model	Link Function		Inverse Link Function	
	Equation	R Code	Equation	R Code
Cumulative Logit	$\text{logit}(p) = \log(\frac{p}{1-p})$	‘qlogis()’	$\frac{e^{\text{logit}(p)}}{1+e^{\text{logit}(p)}}$	‘plogis()’
Cumulative Probit	$z = \Phi(p)$	‘qnorm()’	$\Phi^{-1}(z)$	‘pnorm()’

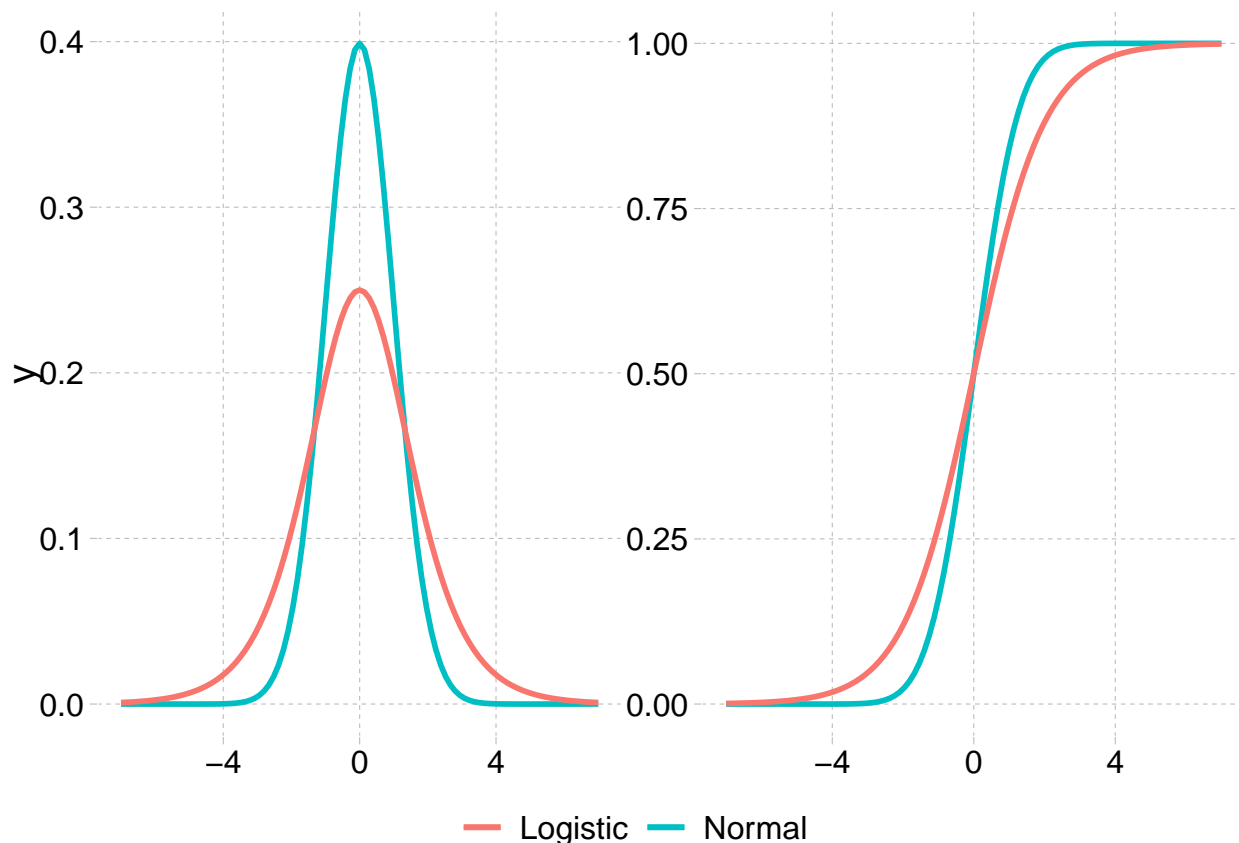


Figure 3. caption here

Model fitting

For fitting the cumulative models we used the `ordinal` package (Christensen, 2019). Despite the presence of other possibilities the `ordinal` package provide the most complete and intuitive way to implement the ordinal models. The syntax is very similar to standard linear models in R and default functions to calculate predictions, perform model comparison, extract relevant model information are implemented similarly to standard regression modelling.¹

The function to fit the model is `c1m()` and the model equation is specified using the

¹ For a very complete overview of the ordinal package see Christensen (2019) and the Package documentation <https://cran.r-project.org/web/packages/ordinal/ordinal.pdf>

R formula syntax $y \sim x$ where y is the ordinal dependent variable and x is one or more predictors eventually including also the interactions. The package also implements mixed-effects models (see the `clmm()` function) including random intercepts and slopes but the topic is beyond the scope of the current tutorial.

When fitting the model the crucial arguments are the `formula`, the `link` function and the `data`. More advanced arguments are the `nominal`, `scale` and `threshold`.

- **formula**: the formula $y \sim x$ with the dependent variable and predictors.
- **link**: is the link function. In this tutorial we consider only the *logit* and *probit* link but other link functions are available.
- **data**: is the dataset where with the variables included in the `formula`
- **nominal**: formula with predictors where the proportional odds assumption (See Section) is relaxed (i.e., partial or non proportional odds)
- **scale**: formula with predictors for the scale (standard deviation) parameter. This argument allow to fit a scale-location model (see Section). The main `formula` argument refers to predictors on the location parameter (i.e., the mean μ).
- **threshold**: different structures for estimating the thresholds. The default is `threshold = "flexible"` where $k - 1$ threshold (where k is the number of ordinal levels for Y) are estimated.

We can start by fitting a simple model, highlighting the crucial parameters where the detailed explanation will be expanded in the next sections. Table contains simulated data from $n = 100$ participants rating the agreement about a certain item with $k = 4$ ordered options. The participants are divided into two groups (predictor x). We can fit a cumulative link model with `clm()` function and check the model summary.

```
## $mean
##      a      b
## 2.52 3.04
```

Table 2
(*ref:tab-dataset-example1*)

id	x	y
1	a	4
2	a	3
3	a	2
4	a	2
...
97	b	3
98	b	3
99	b	2
100	b	4

```
##
## $median
##   a   b
## 2.0 3.5
##
## $sd
##           a           b
## 1.092460 1.124132

##
##           a   b
## 1 0.20 0.14
## 2 0.34 0.18
## 3 0.20 0.18
## 4 0.26 0.50

## formula: y ~ x
## data:    dat
##
## link threshold nobs logLik AIC      niter max.grad cond.H
```

```
## logit flexible 100 -130.61 269.23 5(0) 2.61e-13 2.0e+01
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## xb    0.8795     0.3719   2.365   0.018 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##      Estimate Std. Error z value
## 1|2  -1.2246     0.3040  -4.029
## 2|3   0.1231     0.2653   0.464
## 3|4   0.9396     0.2841   3.308
```

The two main sections of the model summary are the *Coefficients* section reporting the regression coefficients β and the *Threshold* section reporting the α estimation. Given that $k = 4$ we have $k - 1 = 3$ thresholds and a single β associated with the x effect. As in standard regression models, when x is a categorical predictor with j level, we will estimate $j - 1$ regression coefficients where the interpretation depends on the contrast coding (for an overview about contrasts coding schemes see Schad, Vasishth, Hohenstein, & Kliegl, 2020)²

Interpreting parameters

Logit Model. Odds and odds ratio

aggiungere cosa significano i parametri nella regressione

The odds of a probability p is defined as $\frac{p}{1-p}$ thus the probability of success divided by the probability of failure. The odds takes value ranging from 0 to ∞ . For example with a probability of $p = 0.8$ we have an odds of 4, thus we have a 4 successes for each failure. The same as having $p = 0.2$ and an odds of 0.25 means that for each 0.25 successes we

² In R the default is the dummy coding where a factor of j levels is converted into $j - 1$ dummy variables. By default, the first level of the factor is taken as the reference level and the $j - 1$ coefficients represent the comparison between the other levels and the reference.

have a failure or that we have 4 failures for each success. When comparing two groups or conditions we can compare the two odds calculating an odds ratio. The odds ratio is the mostly used statistics to compare groups or conditions on a binary outcome. An odds ratio of 4 means that the odds of success at the numerator is 4 times higher than the odds of success at the denominator. The Figure 4 shows the relationship between probabilities and odds. The logit transformation is about taking the logarithm of the odds creating a symmetric function ranging from $-\infty$ to ∞ with $p = 0.5$ as the midpoint because $\log(\frac{0.5}{1-0.5}) = 0$. The standard logistic regression with two outcome model the logit transformed probability.

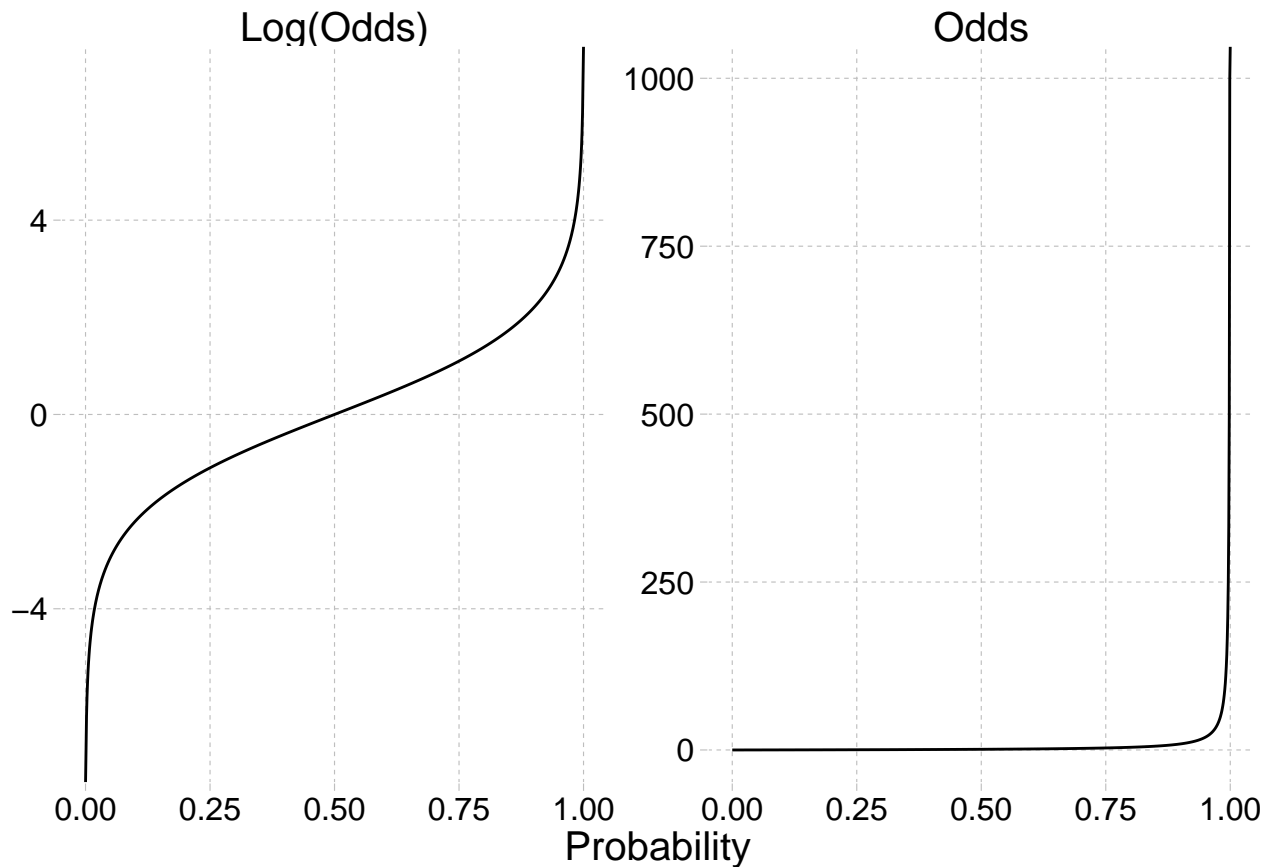


Figure 4. (ref:fig-odds-example)

Odds and odds ratios are clearly defined with $k = 2$ outcomes. With an ordinal variable and a cumulative model we can use the cumulative odds ratio. With e.g. $k = 4$

outcomes we have $k - 1$ models determined by the cumulative probability in terms of $P(Y \leq 1), \dots, P(Y \leq k - 1)$.

Proportional odds assumption

Following again the taxonomy by Gerhard Tutz (2022), each of the presented ordinal regression model has a basic version making the proportional odds assumption. There are more advanced versions of the model relaxing this assumption completely (*non proportional odds* models) and partially (*partial proportional odds* models) (rivedi se la nomenclatura va bene). The proportional odds assumption is a crucial point when interpreting model parameters. Can be formalized as in Equation (). Basically if we use the *logit* link function $g()$ the β s are interpreted as standard odds ratio in the logistic regression. Given that we have $k > 2$ alternatives there is the need of $k - 1$ equations (as reported in Equation ()). The porportional odds is assuming that the regression coefficients β are independent from the thresholds α thus the effect is the same regardless of the Y level. Thus $\beta_1 = \beta_2 \dots = \beta_{k-1}$. The Figure () depicts the proportional odds assumption for the $k - 1$ logistic curves both for probabilities and linear predictors η .

The proportional odds assumption is convenient because regardless the number of k , the β_j (j being the number of regression coefficients) effect is assumed to be the same. The model is more parsimonious compared to estimating $k - 1$ coefficients for each β_j as in the multinomial regression or the non-proportional version of the ordinal regression. Clearly this can be a strong assumption in some conditions. There are several methods to test the proportional odds assumption (see Liu, He, Tu, & Tang, 2023 for an overview). Gerhard Tutz and Berger (2020) suggested that using alternative as the location-shift or location-scale models, the assumption can be made (improving model parsimony) but still having a more flexible modeling framework. In the next section, we provide more details about the location-scale model in terms of parameters interpretation and data simulation.

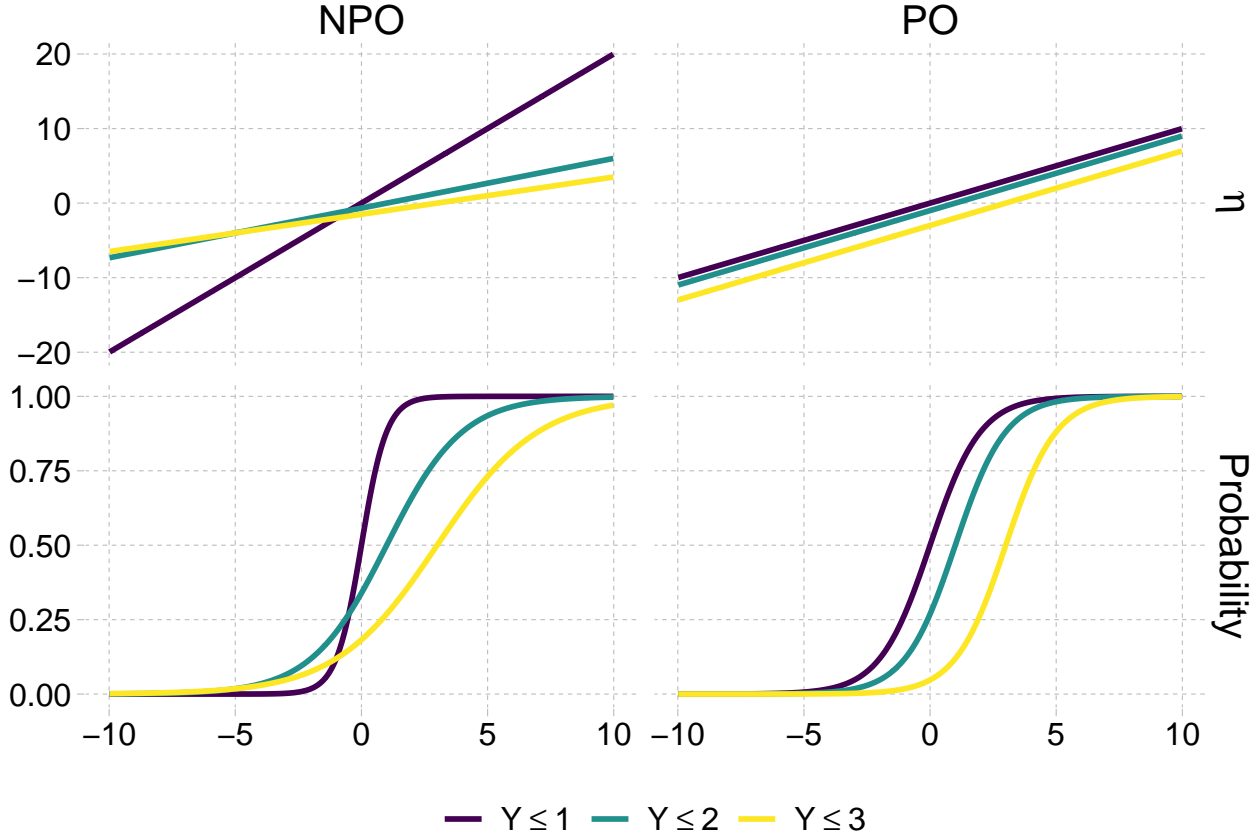


Figure 5. caption here

To make a practical example, we can use again the dataset presented in Table. We can check the proportional odds by extracting fitted probabilities and manually computing the odds ratios.

Equation (4) formalize the proportional odds assumption. Basically the cumulative log odds ratio comparing $P(Y_k|x_0)$ with $P(Y_k|x_1)$ is the same regardless the specific k or threshold α .

$$\text{logit}\left(\frac{P(Y \leq 1|x_1)}{P(Y \leq 1|x_0)}\right) = \dots = \text{logit}\left(\frac{P(Y \leq k-1|x_1)}{P(Y \leq k-1|x_0)}\right) \quad (4)$$

This can be shown by fitting the model with `c1m()` that by default assume the proportional odds and using the predicted probabilities to compute the odds ratios. In the following code we computed the odds ratio comparing x_a and x_b when $k \leq 1$ and $k \leq 2$.

The log odds ratio is the same.

```
# fitting the model
fit <- clm(y ~ x, data = dat, link = "logit")

# extracting the predicted probabilities for the two groups
pr <- predict(fit, data.frame(x = unique(x)))$fit

odds <- function(p) p / (1 - p) # function to compute odds

# group a
a_or1vs2345 <- odds(pr[1, 1]) # 1 vs 2 3 4 5
a_or12vs345 <- odds(sum(pr[1, 1:2])) # 1 vs 2 3 4 5

# group b
b_or1vs2345 <- odds(pr[2, 1]) # 1 vs 2 3 4 5
b_or12vs345 <- odds(sum(pr[2, 1:2])) # 1 vs 2 3 4 5

c(xa = log(a_or1vs2345 / a_or12vs345), xb = log(b_or1vs2345 / b_or12vs345))

##          xa          xb
## -1.347649 -1.347649
```

Clearly, the odds ratio here is exactly the same because the model assume the proportional odds assumption putting this constraint on the predicted probabilities. The estimated β for each $k - 1$ equation could differ and the amount of heterogeneity is actually a way to test the proportional odds with real data. In the supplementary materials we implemented an intuitive plot proposed by Harrell (2015) to visually assess the proportional odds assumption.

Proportional odds and *probit* model. The proportional odds assumption is relevant only for the *logit* model because of the link function that put β as log odds. However the concept of the proportional odds that is more generally called parallel slopes assumption is assumed also for the probit model. Equation (??) depict the parallel slopes assumption for a probit model. Basically the difference in z for a unit increase in x (i.e., the slope) is the same regardless the Y level. This can be directly demonstrated in the

same way as for the logit model. We use the same dataset presented in Table ?? but fitting a probit model. The difference between the z scores comparing the two groups calculated on the cumulative probability is the same regardless the $k - 1$ level.

$$z_{x_1-x_0} = \Phi(P \leq k|x_0) - \Phi(P \leq k|x_1) = \dots = \Phi(P \leq k-1|x_0) - \Phi(P \leq k-1|x_1)$$

```
# fitting the model
fit <- clm(y ~ x, data = dat, link = "probit")

# extracting the predicted probabilities for the two groups
pr <- predict(fit, data.frame(x = unique(x)))$fit

# group a
a_1vs2345 <- qnorm(pr[1, 1]) # 1 vs 2 3 4 5
a_12vs345 <- qnorm(sum(pr[1, 1:2])) # 1 2 vs 3 4 5

# group b
b_1vs2345 <- qnorm(pr[2, 1]) # 1 vs 2 3 4 5
b_12vs345 <- qnorm(sum(pr[2, 1:2])) # 1 2 vs 3 4 5

c(xa = a_1vs2345 - a_12vs345, xb = b_1vs2345 - b_12vs345)

##          xa          xb
## -0.7963604 -0.7963604
```

Beyond the proportional odds. qualcosa che spieghi come andare oltre e qualche proposta per testare l'assunzione.

Probit Model z scores

The alternative to fitting a cumulative logit model is using a standard normal distribution thus fitting a *probit* model. The main difference is that the latent distribution is now a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$ and the link function is the cumulative gaussian distribution function Φ implemented in R with the `pnorm()` function.

The inverse of the Gaussian cumulative distribution function is Φ^{-1} implemented in R with the `qnorm()` function. The logistic and normal distribution are very similar, with the logistic having more variance.

The main difference regards the parameters interpretation. In the logit model the β are the log odds ratio. For categorical variables they represents the increase in the log odds of moving from one level to the other while for numerical variables is the increase in the log odds for a unit increase in x . For the probit model, the β is the increase in terms of z scores for a unit increase in x . This is very convenient especially for categorical variables because parameters can be interpreted as a Cohen's d like measure. Thinking about the latent distributions, the β is the shift in the latent mean comparing two or more groups or the slope of latent scores as a function of a numeric x . The interpretation in terms of shifting the latent mean holds also for the logistic model. However, the standard deviation of the standard logistic regression is $\frac{\pi}{\sqrt{3}} \approx 1.81^3$. The β for the logistic distribution can be interpreted as the location shift of the latent logistic distribution by $\frac{\pi}{\sqrt{3}}$ standard deviations.

Another advantage of the probit model is the possibility to directly map parameters from signal detection theory (Green, Swets, & Others, 1966; Stanislaw & Todorov, 1999) into the model coefficients. In practical terms, the thresholds α are the criterion cut-offs and the β_1 is the d' parameter (DeCarlo, 1998; Knoblauch & Maloney, 2012).

qualcosa in più qui

³ Actually the variance of the logistic distribution is $\frac{s^2\pi^2}{3}$ and the standard deviation $\frac{s\pi}{\sqrt{3}}$ where s is the scale of the distribution. For the standard logistic distribution $s = 1$ (as for the standard normal distribution). For the standard normal distribution $s^2 = \sigma^2 = 1$

Simulating data

There are mainly two ways to simulate data. The first method concerns simulating starting from the latent formulation of the ordinal model. Basically we can simulate the underlying latent distribution and then fixing the thresholds converting the latent continuous variable into the observed ordinal variable.

Simulating from a multinomial distribution. The first method to simulate ordinal data as a function of predictors is by calculating the true probabilities $g^{-1}(\eta)$ as a function of predictors and then sample from a multinomial distribution⁴. This is similar to the general method to simulate data for a generalized linear model where the linear predictors η are calculated and data are generated from the corresponding probability distribution of the random component. The following code box simulate a n binary trials with a continuous predictor x .

Essentially the `lp` (η_i) vector contains true probability of success ($p_i|x_i = g^{-1}(\eta_i)$) for the x_i level of the predictor. Then to introduce the random part of the model, we can use p_i to generate binary values from a Binomial (in this specific case a Bernoulli) distribution. This create `y` that is the vector of 0-1 values generated using the true probabilities of success. The supplementary materials contains a more extensive example for simulating data for generalized linear models.

We can apply the same idea to an ordinal outcome but we need $k - 1$ equations (compared to the previous example) where k is the number of ordinal levels. Let's simulate a similar design with a continuous x predictor and $k = 4$ options. We fix the baseline probabilities where $x = 0$ as uniform thus $p(y_1) = p(y_2) = \dots p(y_k) = 1/k$. The proposed workflow for the simulation is:

⁴ In R we are using the `sample()` function that simulate values from a *categorical* distribution. The *categorical* distribution is considered a special case of the multinomial distribution

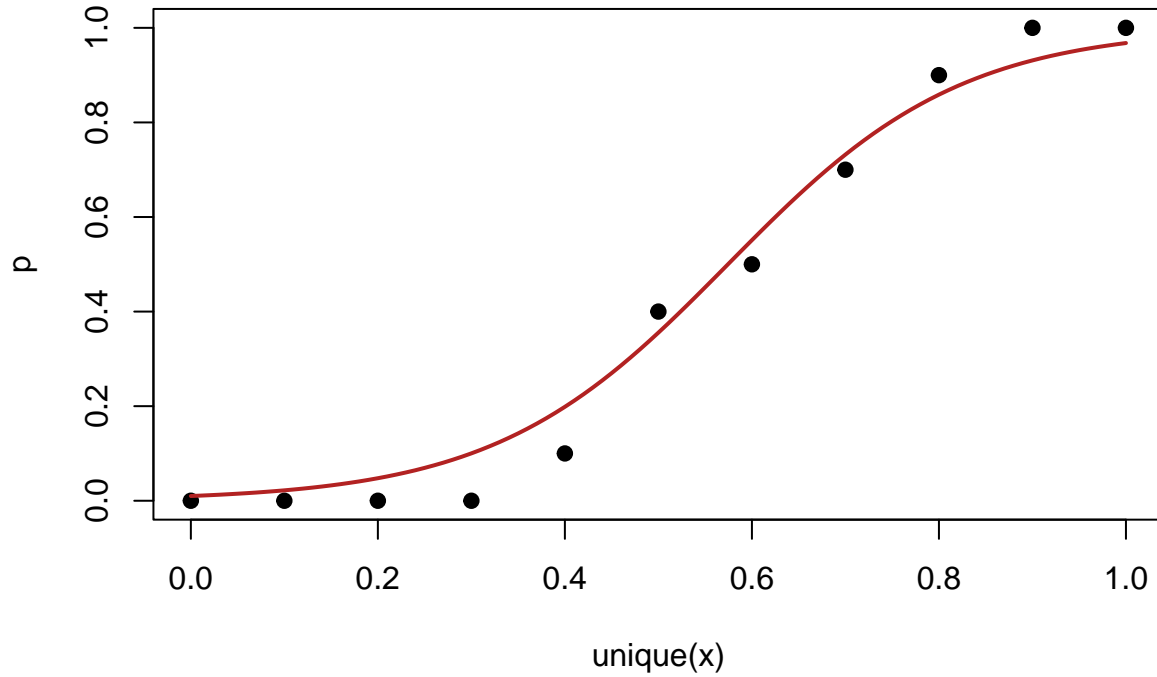


Figure 6

1. Define n (number of observations) and k (number of ordinal outcomes)
2. Define the regression coefficients β s.
3. Define the baseline probabilities when $x = 0$. These probabilities can be converted to the corresponding thresholds α choosing the appropriate link function (*logit* or *probit*).
4. Calculate the linear predictors η using $k - 1$ equations combining the predictors \mathbf{X} and regression coefficients β .
5. Apply the inverse of the link function $g^{-1}(\eta)$ on the linear predictor and calculate the cumulative probabilities $p(y \leq 1|x), p(y \leq 2|x), \dots p(y \leq k - 1|x)$.
6. Calculate for each observation the probability of the k outcome as the difference between cumulative probabilities. For example, the probability of

$P(y = 2) = P(y \leq 2) - P(y \leq 1)$. This is implemented in R by adding a columns of zeros and ones respectively at the beginning and end of the matrix of cumulative probabilities because the probability of the first $Y = 1$ level is the cumulative probability from $-\infty$ to the first threshold and the probability of $Y = k$ is obtained removing $P(Y \leq k - 1)$ from one.

7. Sample n outcomes from a multinomial distribution, choosing between k alternatives with the calculated probabilities.

The linear predictors (η) can be seen in the Figure ref. The g in this case is the cumulative *probit* function Φ .

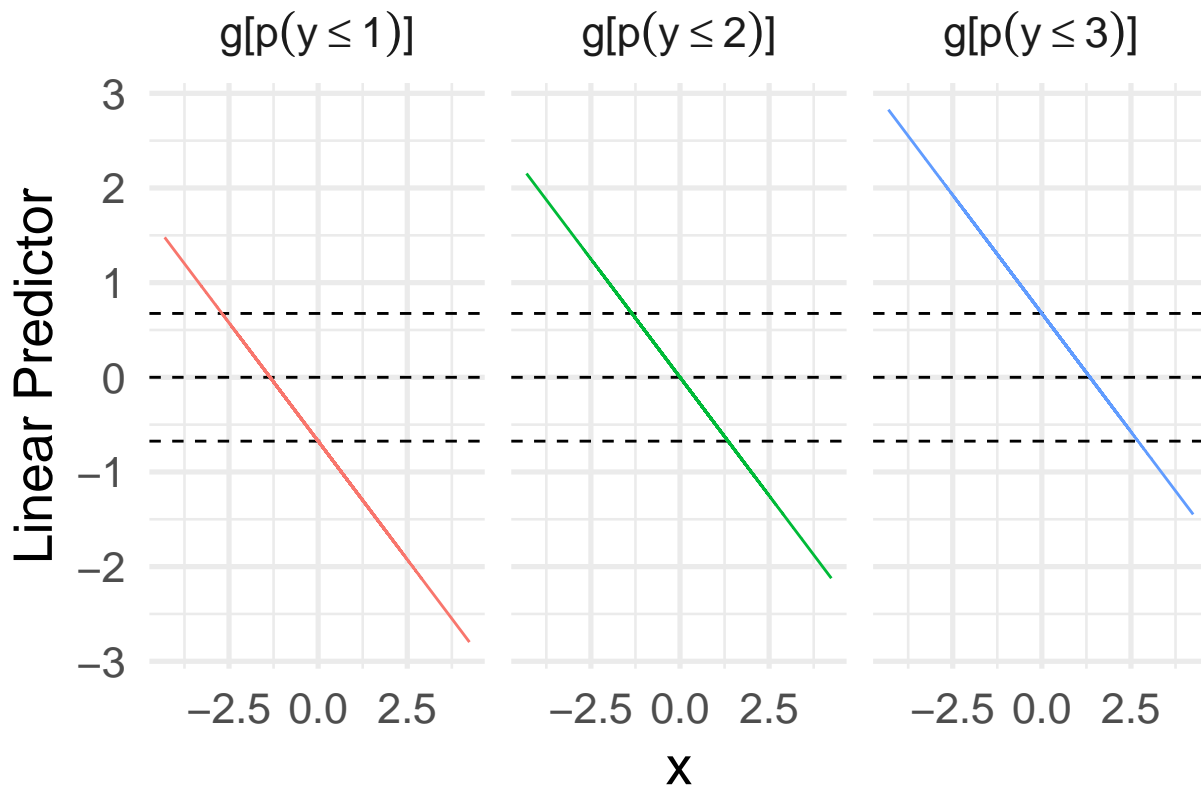


Figure 7

Now we can apply the inverse of the link function $g^{-1} = \Phi^{-1}$ to calculate the corresponding cumulative probabilities.

```
##   p_y1 p_y2 p_y3 p_y4
## 1 0.088 0.161 0.25 0.502
## 2 0.296 0.259 0.237 0.208
## 3 0.265 0.254 0.246 0.235
## 4 0.239 0.247 0.253 0.261
## 5   ...   ...   ...   ...
## 6 0.143 0.205 0.264 0.389
## 7 0.167 0.218 0.264 0.35
## 8 0.338 0.263 0.223 0.176
## 9 0.265 0.253 0.246 0.236
```

We can plot the expected effect of β_1 on the k probabilities using the `num_latent_plot()` function (see Figure 8). Finally we sample using the `sample()` function using the calculated probabilities.

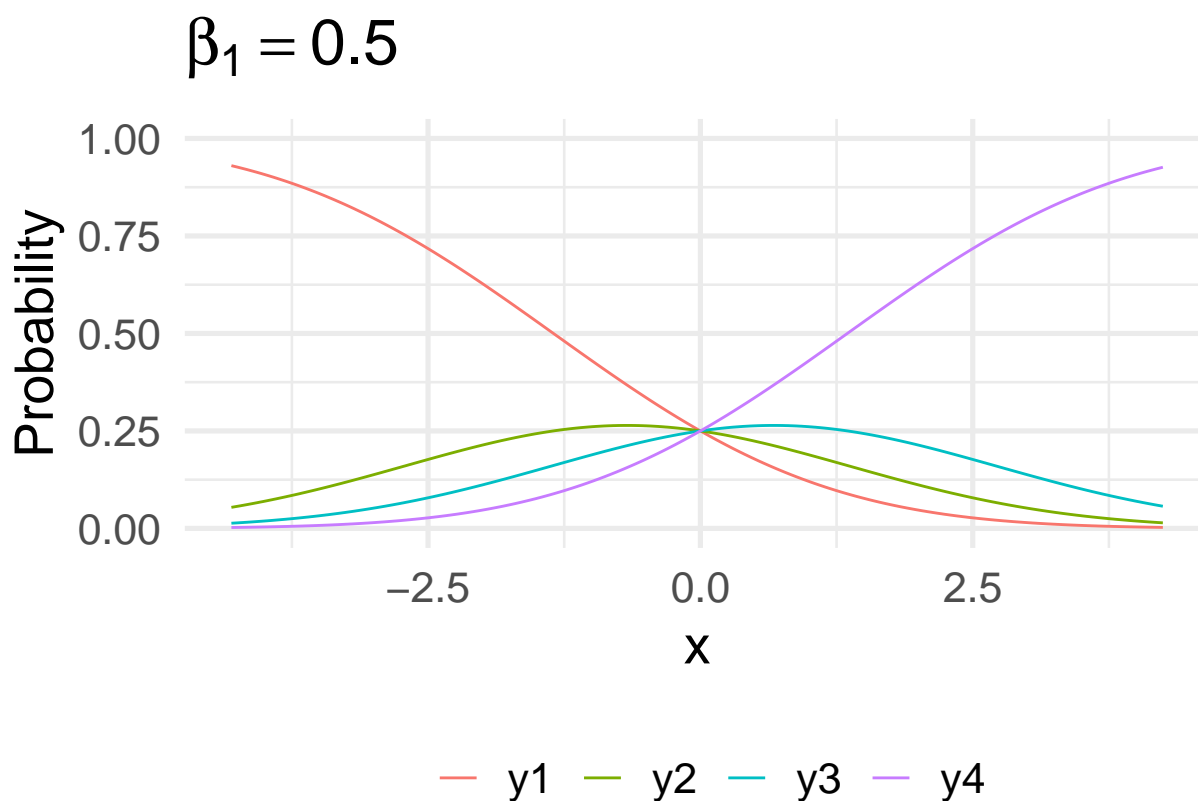


Figure 8. sim-with-probs-lat-plot

To check the simulation result we can increase the number of observations, fit the

model using the `ordinal::clm()` function and assess the recovery of simulated parameters.

Simulating from the latent distribution. An equivalent but more efficient way to simulate an ordinal outcome is using the latent formulation of the model. This requires simulating a standard linear regression using the appropriate data generation function (*logistic* or *normal*) and then cutting the latent values according to the thresholds α . The workflow is slightly different compared to the previous approach.

1. Define n (number of observations) and k (number of ordinal outcomes)
2. Define the regression coefficients β
3. Define the baseline probabilities when $x = 0$. These probabilities can be converted to the corresponding thresholds α choosing the appropriate link function (*logit* or *probit*).
4. Calculate the linear predictors η using $k - 1$ equations combining the predictors \mathbf{X} , the regression coefficients β and the error sampled from the appropriate probability distribution.
5. Cut the latent variable into k areas using the thresholds α and assign the corresponding ordinal value. This step simply checks if the latent value Y_i^* is between two threshold values and assign the corresponding value. This can be done using the `cut()` or the `findInterval()` functions.

The Figure (put crossref) depicts the simulated Y^* and the corresponding ordinal value. As for the previous simulation we can fit the model using `clm()` and check the estimated parameters.

The simulation using the latent formulation of the model is implemented in the `sim_ord_latent()` function.

Choosing parameters values.

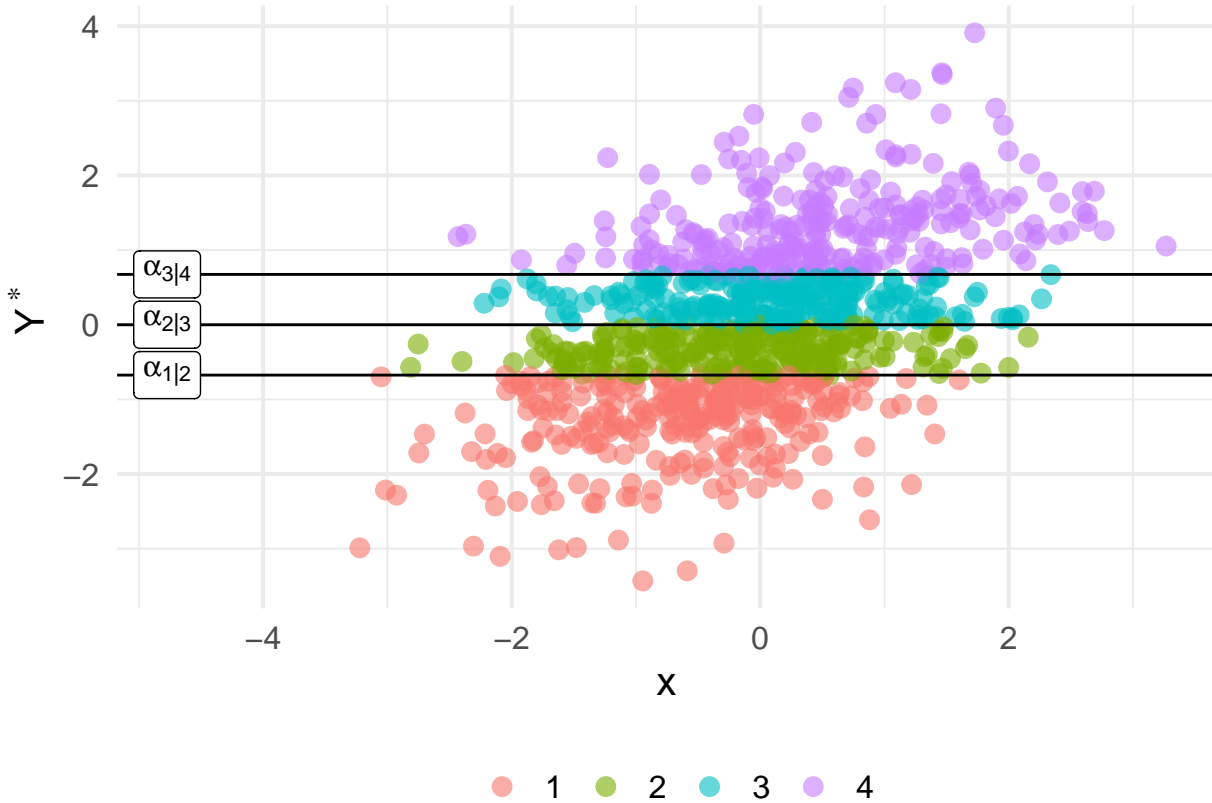


Figure 9

Thresholds α . The two proposed simulation workflows can be really flexible for simulating complex models with multiple predictors and interactions. The critical part of the simulation is setting appropriate and empirically meaningful parameters values. The thresholds are usually not the crucial part of the statistical inference and model interpretations. In fact, they can be considered as the intercept in standard linear models. In any case to set meaningful α values the best strategy is to convert them into the corresponding probabilities according to the link function. The thresholds are essentially quantiles of the latent distribution (logistic or normal) without a direct empirical interpretation. The distance between the thresholds determine the probability assigned to a specific Y value. The function `show_alpha()` allow to produce a meaningful plot starting from the thresholds or the probabilities. For example, the Figure 10 depict the

threshold/probabilities associated with a four-level ordinal variable.

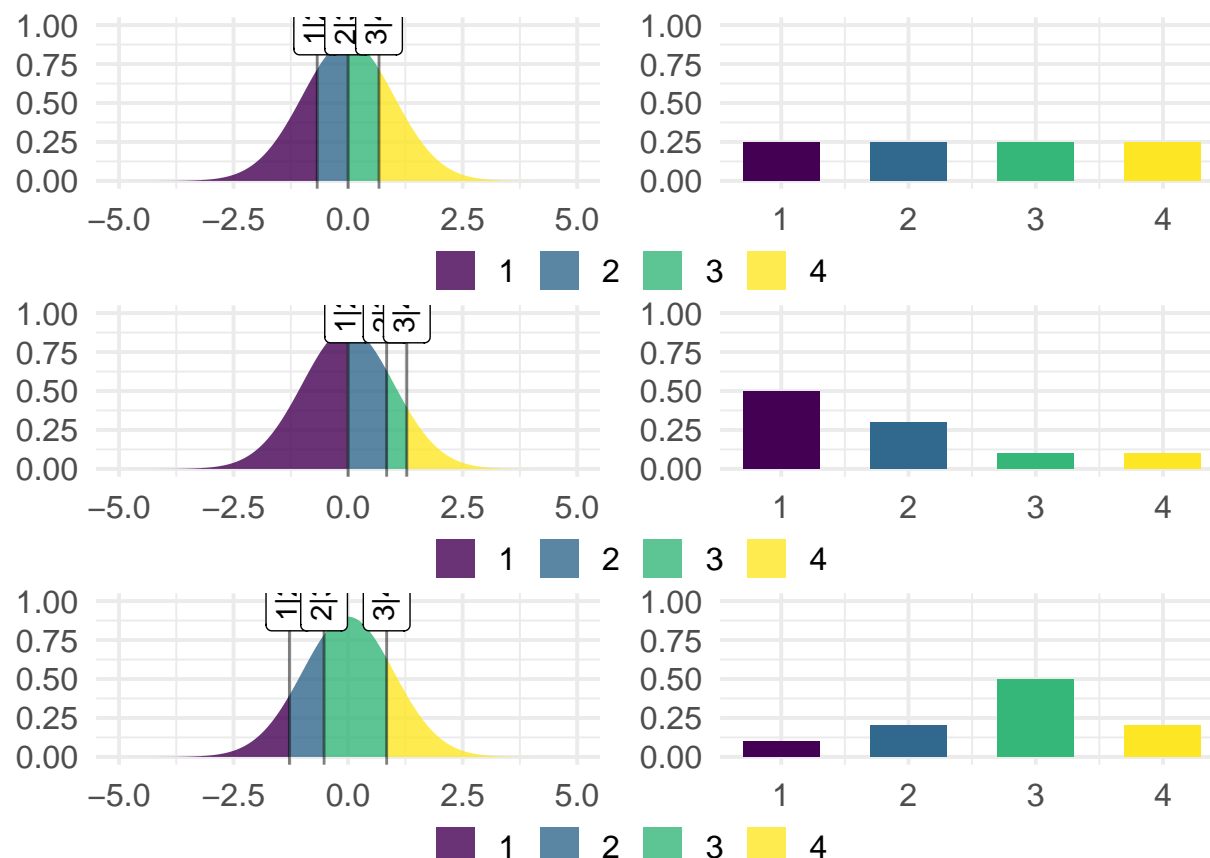


Figure 10

Thus the thresholds can be fixed according to the desired probabilities distribution. For example, when simulating the difference between two groups on an ordinal variable, the threshold can be set thinking about the response distribution in the reference group. The `alpha_to_prob()` and `prob_to_alpha()` functions can be used to convert between probabilities and thresholds.

```
## [1] -0.2533471  0.0000000  0.8416212
## [1] 0.02275013 0.47724987 0.34134475 0.15865525
```

Regression coefficients β . As reported in the sections above, the regression coefficients have a different interpretation according to the link function. Fortunately, both parameters have an intuitive interpretation given that they are directly expressed in a

interpretable scale. For probit models we can think about standardized mean differences (e.g., Cohen's d) between categorical variables or standardized regression coefficient for a numerical predictor. For logit models we can fix the parameters thinking about the odds ratios. Meaningful odds ratios can be derived from previous literature, meta-analyses or converting to other effect sizes. For example, Sánchez-Meca, Marín-Martínez, and Chacón-Moscoso (2003) proposed some equations to convert between odds ratios and Cohen's d . Using their approach, a Cohen's d of 0.5 usually considered a plausible medium effect size corresponds to an odds ratio of ≈ 2.47 . Another more informative strategy is about plotting the predicted probabilities $g^{-1}(\eta)$ associated with certain regression coefficients. The `cat_latent_plot()` and `num_latent_plot()` functions can be used to see the impact of choosing a specific regression coefficients for respectively a categorical (Figures 11) and numerical predictor (Figures 12).

```
cat_latent_plot(m = c(0, 0.5), s = 1, probs = rep(1/4, 4), link = "logit", plot = "both")
```

```
num_latent_plot(x = runif(100), b1 = 2, probs = c(0.6, 0.2, 0.1, 0.1), link = "probit")
```

Scale Effects

The default ordinal regression model assume that the variance of the underlying latent distribution is the same across condition. This is similar to a standard linear regression assuming the homogeneity of variance. For example, when comparing two groups or conditions we can run a standard linear model (i.e., a t-test) assuming homogeneity of variances or using the Welch t-test (see Delacre, Lakens, & Leys, 2017). In addition, there are the so-called location-scale models that allows to include predictors also for the scale (e.g., the variance) of the distribution. This can be done also in ordinal regression where instead of assuming the same variance between conditions, the linear predictors can be included. The Equation (number here) expand the previous model including the linear predictor on the scale of the latent distribution.

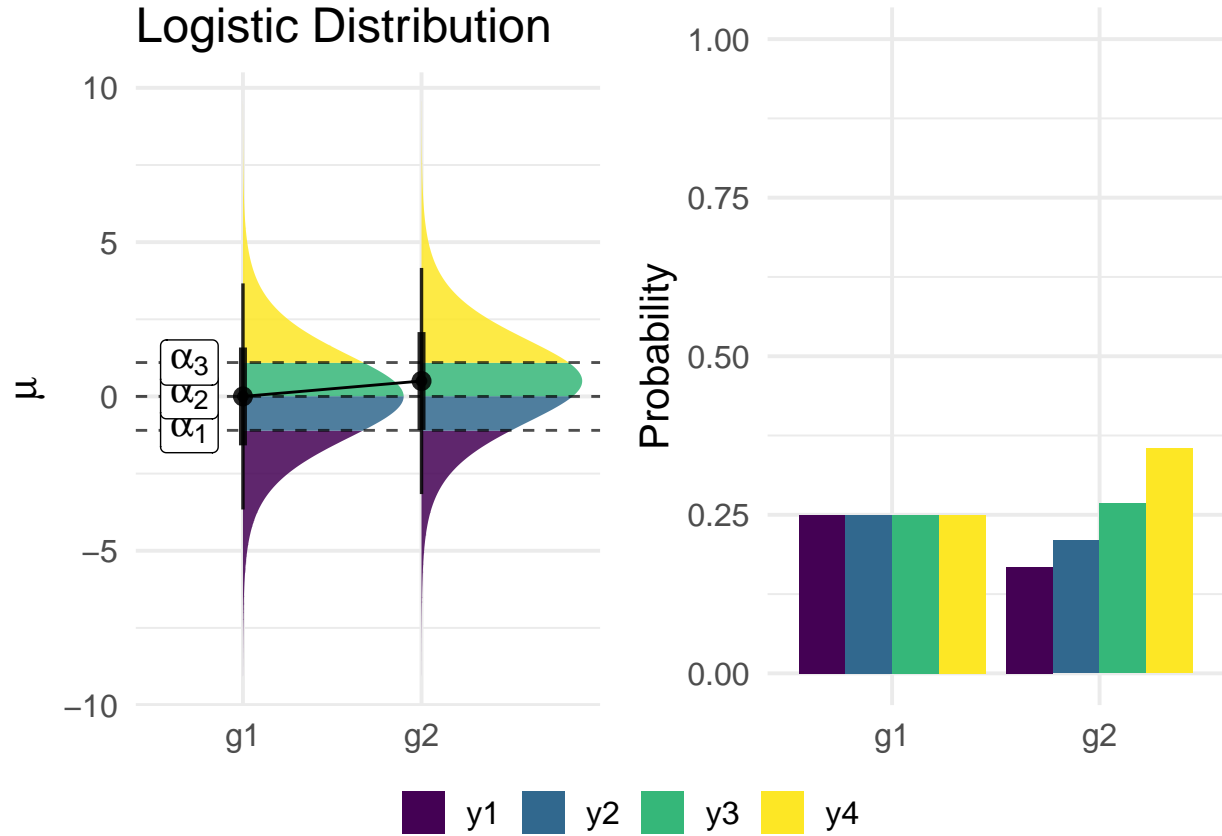


Figure 11. caption here

$$g(P(Y \leq k)) = \frac{\alpha_k - X\beta}{e^{X\zeta}} \quad (5)$$

$$Y_i^* = \eta + \epsilon_i \epsilon_i \sim \mathcal{N}(0, e^{X\zeta}) \quad (6)$$

Where $X\zeta$ is the linear predictor η for the scale of the distribution. By default for both the logit and probit model the scale is fixed to 1. On scale-location models we put predictors on both parameters. Given that the scale cannot be negative we use a log link function $\eta = \log(X\zeta)$. Figure ?? depict an example of a comparison between two groups where the two underlying latent distributions have unequal variance. Ignoring the heterogeneity of variances could lead to underestimation of the effect size, reduced power and inflated type-1 error rate (qualche ref qui). Furthermore, as suggested by G. Tutz and

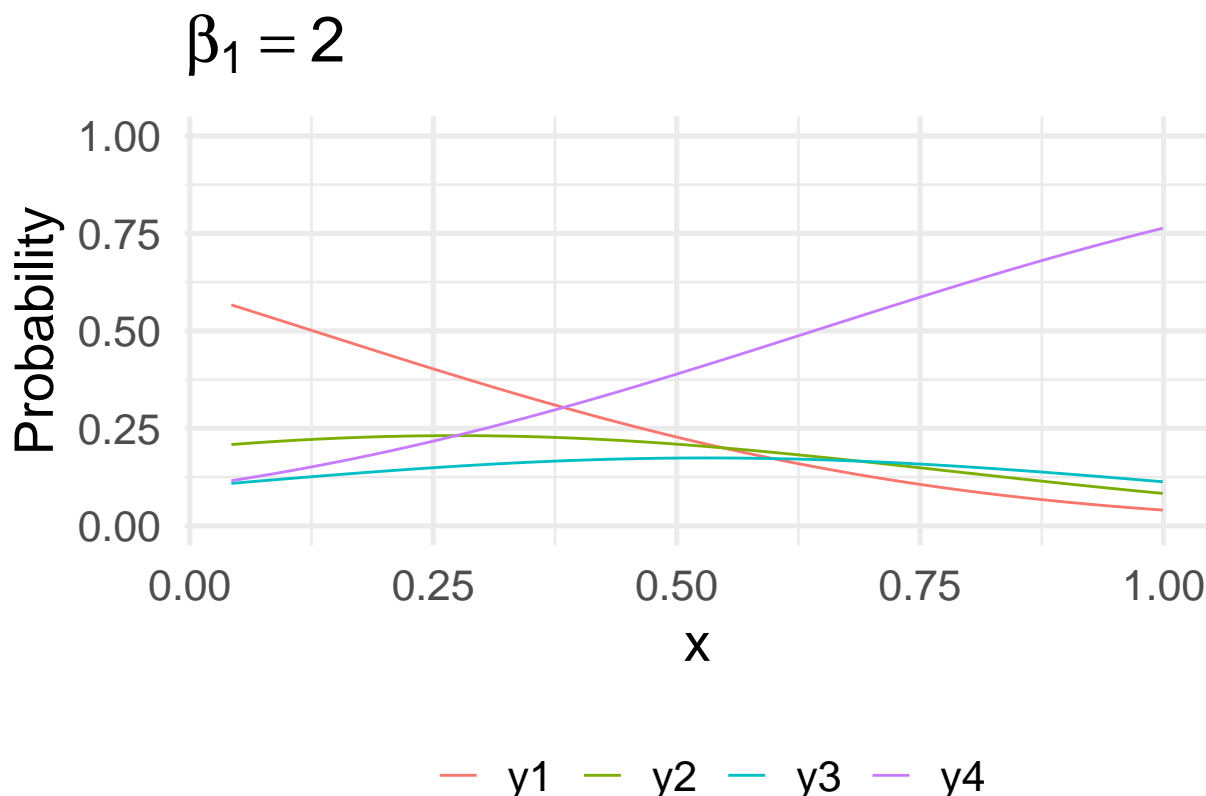


Figure 12. caption here

Berger (2017) location-scale models can be considered as a more parsimonious approach compared to partially or completely relaxing the proportional odds assumption. Allowing the scale to be different as a function of the predictor create more modelling flexibility. Furthermore, two groups could be theoretically different only in the scale of the latent distribution with a similar location. In this example, the only way to capture group differences is by including a scale effect. The Figure ?? depict the impact of having different scales between two groups on the ordinal probabilities.

Gerhard Tutz (2022) provide a very clear and intuitive explanation of what happen when including a scale effect and how to interpret the result. As suggested before, the scale-location model allow to independently predict changes in the location and the scale. While location shifts are simply interpreted as increasing/decreasing the latent μ or the

odds of responding a certain category scale effects are not straightforward. As the scale increase (e.g., the variance increase) there is an higher probability mass on extreme categories. On the other side as the scale decrease, responses are more concentrated on single categories. The categories are determined by the location parameter. For example, if one group have a certain latent mean μ_1 and a small scale $\sigma^2 = 1/3$ (thus one third compared to the standard version of the distribution), all responses will be focused on categories around the latent mean. On the other side, increasing the scale will increase the cumulative probabilities for all categories and for values that tends to infinity extreme categories are preferred. Clearly, the scale parameter can somehow be interpreted as the response style (concentrated or variable). This is conceptually similar but implemented and interpreted differently to shift-location models [Gerhard Tutz (2022); Tutz2020-xq] where thresholds are allowed to vary as a function of predictors. Both models tries to increase the flexibility of modelling the probability structure of the responses by adding extra parameters to predict the probabilities of the ordinal responses.

Simulating scale effects. The location-scale model can be simulated using the `sim_ord_latent()` function and providing the predictors for the scale parameters. Given the `log` link function, predictors are provided on the `log` scale. For example, we simulate the effect of a binary variable x representing two independent groups predicting the $k = 5$ response. We simulate a *location* effect of $\beta_1 = 0.5$ (in *probit* scale) and $\zeta_1 = \log(2) = 0.70$. The first group has a $\sigma = 1$ and the second group has $\sigma = 2$. Again we simulate that the baseline probabilities are uniform for the first group.

The table (put ref) reports the simulation results.

To better understand the impact of assuming (or simulating) a different latent scale we fit $k - 1$ binomial regressions and check the estimated coefficients. We are not simulating a specific beta for each outcome but simulating a scale effect is actually impacting the regression coefficients. When generating data for a binary outcome the linear predictor is composed by $\eta = \beta_0 + \beta_1 x$. The threshold α and slope of the function can be

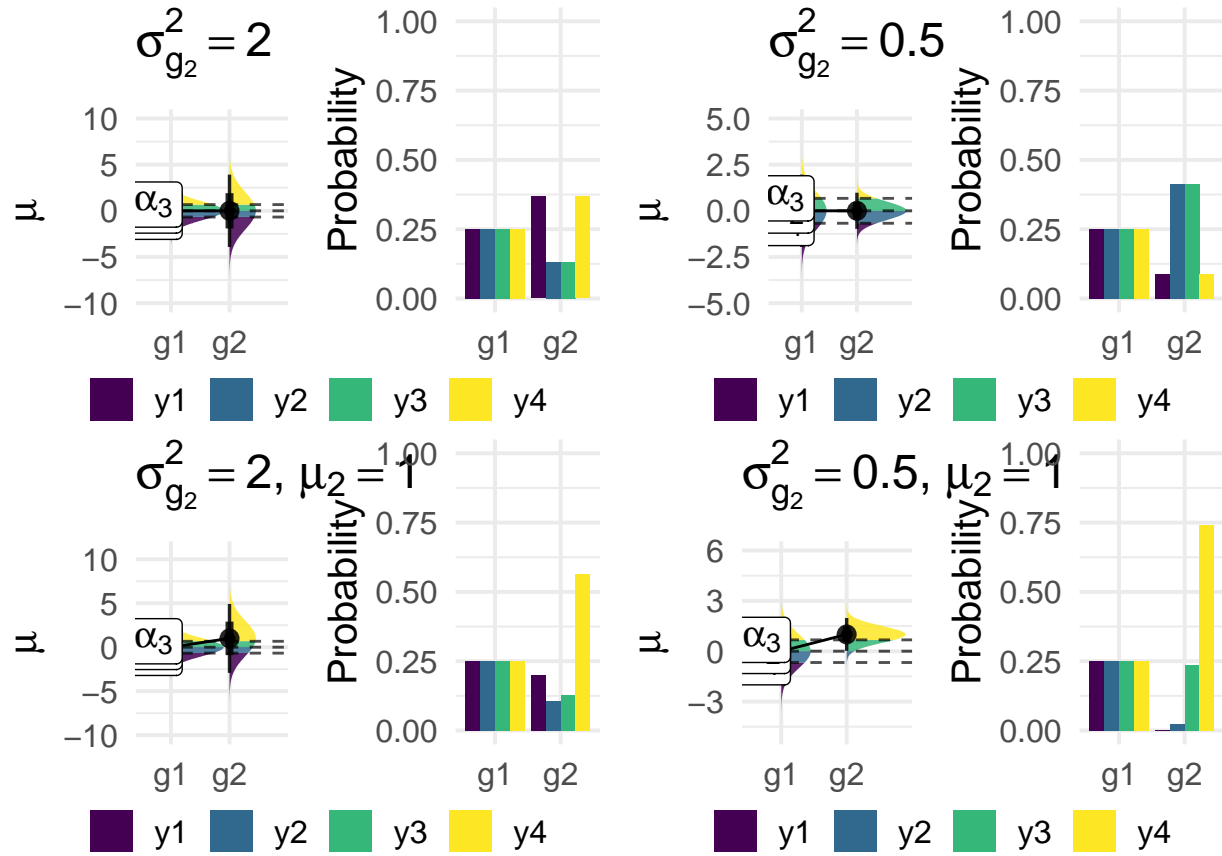


Figure 13

estimated using $\alpha = -\frac{\beta_0}{\beta_1}$ and the slope is $\frac{1}{\beta_1}$ (Faraggi, Izikson, & Reiser, 2003; Knoblauch & Maloney, 2012). Under the proportional odds assumption, there is only a change in thresholds α this a shift in the sigmoid along the x axis. When including a scale effect a change in the sigmoid is combined with a change in the slope.

2x2 interaction

A common research design could be a 2x2 factorial design. In this example we have 2 main effects and the interaction. Regardless the link function, we can focus the model equation on the linear predictor. Equation depicts the linear predictor for a 2x2 interaction. With a 2x2 design, a good strategy is to use sum to zero contrasts for the two factors. By default R use dummy coding but here we set the contrasts for X_1 and X_2 as 0.5 and -0.5 .

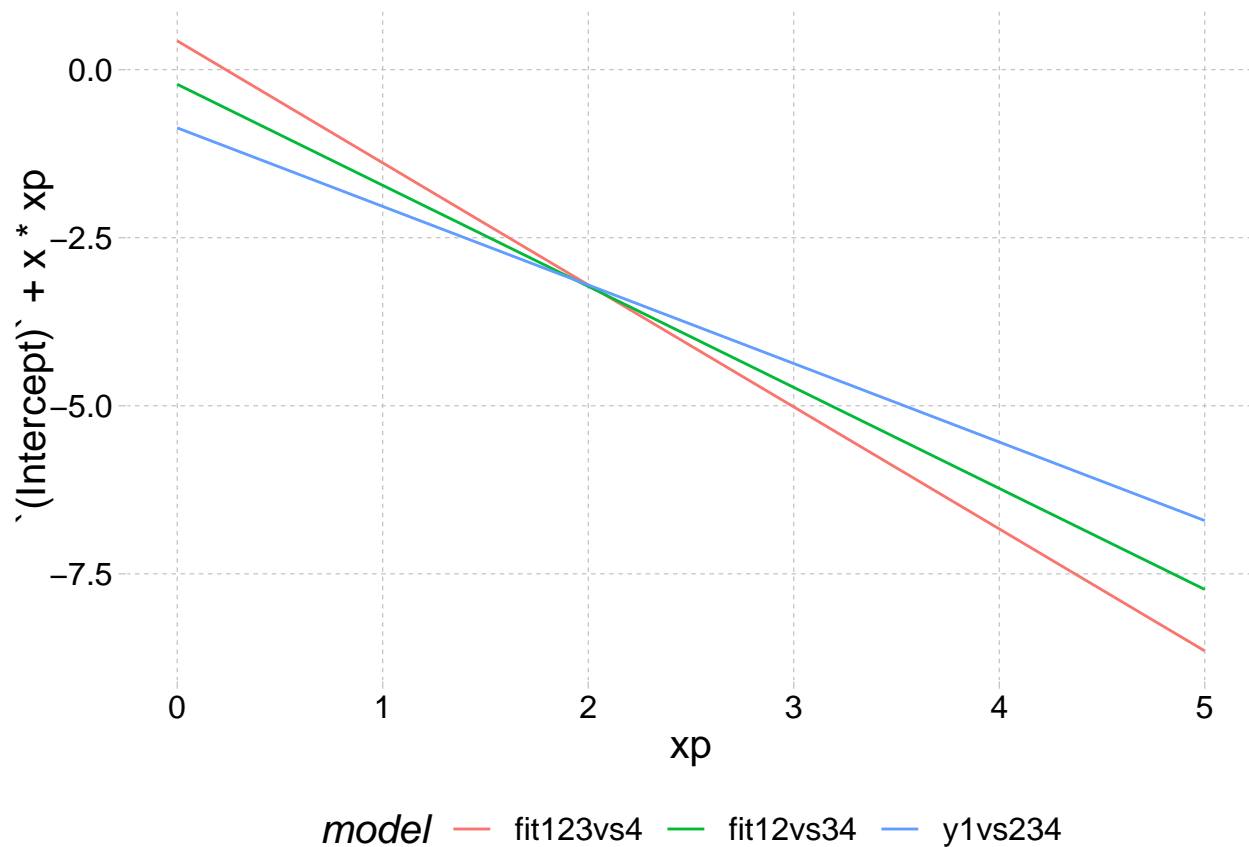


Figure 14

In this way β_1 will be the main effect of X_1 , β_2 the main effect of X_2 and β_3 the interaction (thus the difference of differences).

$$g(P(Y \leq k))^{-1} = \alpha_k - \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{1_i} X_{2_i}$$

```
## formula: y ~ x1 * x2
## data:    dat
##
## link threshold nobs logLik AIC niter max.grad cond.H
## probit flexible 4e+05 -597389.03 1194792.06 5(0) 1.42e-07 4.0e+01
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## x11 0.004798 0.003410 1.407 0.159
## x21 1.004223 0.003519 285.383 <2e-16 ***
```

```
## x1:x21 0.505333 0.006833 73.954 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##      Estimate Std. Error z value
## 1|2 -0.841241 0.002332 -360.8
## 2|3 -0.254088 0.002092 -121.4
## 3|4 0.250954 0.002092 120.0
## 4|5 0.839533 0.002331 360.2
```

The parameters interpretation is the same as introduced in the general case. The thresholds α are fixed the determines the probabilities $P(Y = k)$ when all predictors are zero. In this case by doing `alpha_to_prob(fit$alpha, link = "probit")` we should recover the `bprobs` vector. `x1` is the main effect thus the difference in z scores between a and b averaging over `x2`. The same holds for `x2`. `x1:x2` is the difference of differences in z scores thus $(ad - ac) - (bd - bc)$.

To visualize the effects there is the `ggeffect::ggpredict()` function that compute the predictions according to the specified predictors combination. In this case we can plot the interaction using the estimated probability for each group and Y response.

Numerical by factor interaction

Another common scenario is to compute and interaction between a numerical variable x and a factor g . For simplicity we simulate the factor with two levels and the numerical variable sampled from an uniform distribution $x \sim \mathcal{U}(0, 1)$. Again, the thresholds α represents the baseline probabilities when $x = 0$ and $g = 0$.

```
## formula: y ~ x * g
## data:    dat
##
## link threshold nobs logLik AIC niter max.grad cond.H
## probit flexible 1e+05 -153256.51 306527.02 5(0) 7.97e-08 1.6e+02
```

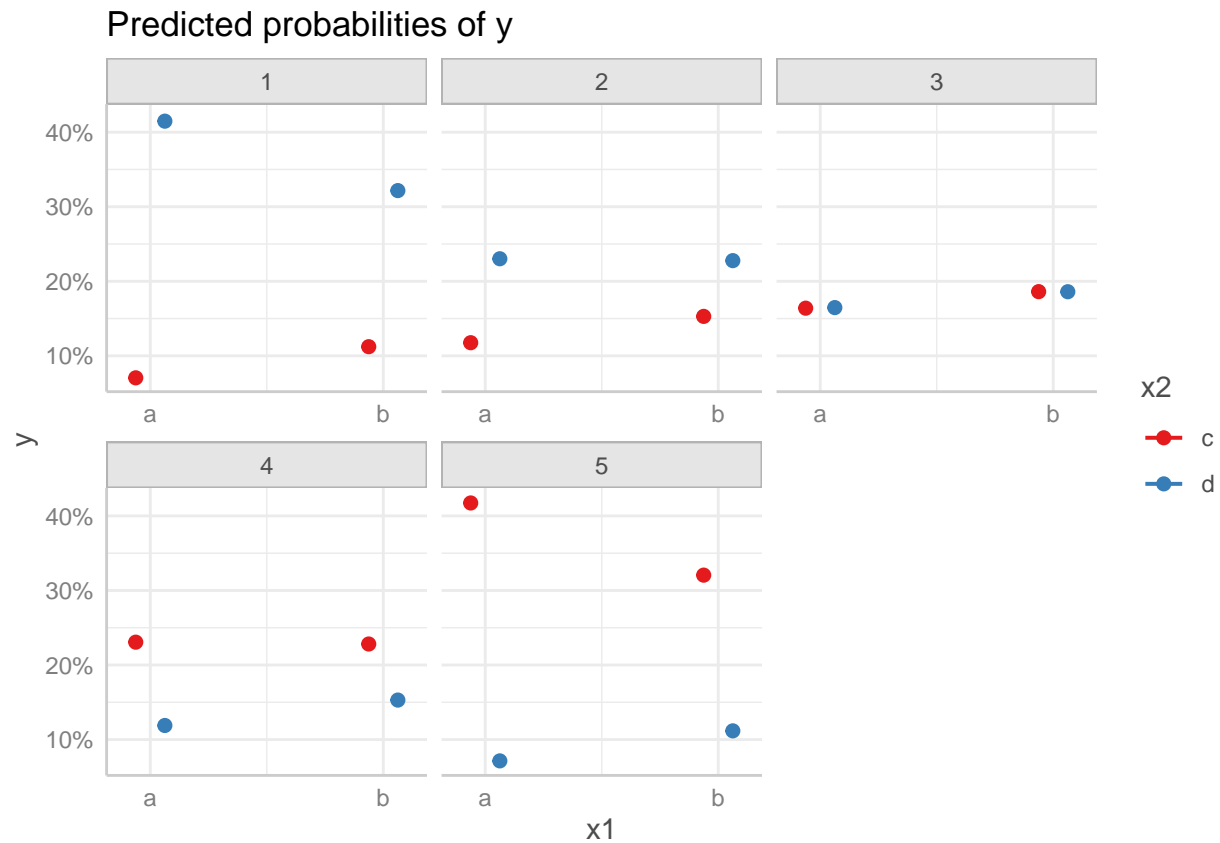


Figure 15

```
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## x      0.29419   0.01172  25.11  <2e-16 ***
## g1     0.49679   0.01351  36.78  <2e-16 ***
## x:g1    0.51690   0.02343  22.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##      Estimate Std. Error z value
## 1|2 -0.841442   0.007474 -112.59
## 2|3 -0.258062   0.007133  -36.18
## 3|4  0.250592   0.007124   35.18
## 4|5  0.838293   0.007404  113.23
```

Again, x is the slope of the numerical predictor averaging over g (given that g has

been coded with sum to zero contrasts) thus the increase in z scores for a unit increase in \mathbf{x} . $\mathbf{g1}$ is the main effect of the factor evaluated when $x = 0$. To change the interpretation of $\mathbf{g1}$ a possibility is to center \mathbf{x} differently (e.g., mean centering). The $\mathbf{x:g1}$ is the difference in slopes between the two groups.

Regardless of the model complexity, the `sim_ord_latent()` function can be expanded by including more predictors. Clearly, as the model complexity increase, choosing meaningful and empirically plausible parameters became more challenging.

Power Analysis

In the previous sections we introduced how to simulate standard and scale-location cumulative models and how to fit the corresponding models using the `clm()` function from the `ordinal` package. In this section we introduce how to compute the power for some common scenarios. We present an example in a very general form that can be expanded for specific use cases. For example, by simulating a model fixing all effects to zero it is possible to estimate the type-1 error rate or fixing the sample size and varying the effect size it is possible to do a sensitivity analysis focusing on the smallest effect size of interest (ref qui). The general workflow for a power simulation can be summarized as:

1. Choose the research design, e.g., 2x2 factorial design
2. Indentify the specific effect to focus on. For example the interaction effect, a main effect or a specific contrast
3. Write the model equation similarly to what presented in the previous sections and extract the model coefficients
4. Choose one of the proposed strategy for simulating data and follows the corresponding steps
5. Fit the model and extract the relevant statistics e.g., the p-value of the interaction

6. Repeat the simulation a large number of times (e.g, 10000) and store the relevant values from each simulation
7. Summarise the simulation results in the appropriate way. For example, the power of the interaction is the number of times the p-value for the interaction test is lower than the critical level divided by the number of simulations.

To make a practical example, we simulate the power of a cumulative link model in detecting an effect of $d = 0.4$ (probit model) with $k = 6$ and $n = 40$ participants per group. We fix the baseline probabilities in the first group as uniform.

descrizione migliore

Then to compute the power we can simply compute the proportion of p values lower than the critical level. This can be done in R as `mean(p <= alpha)` resulting in an estimated power of 0.39. Clearly using only a single condition is not really informative thus we can vary the sample size across plausible values and check the corresponding power curve.

Disclaimer about the functions

The current paper proposed a simplified way with some functions to generate ordinal data. For more complex simulations such as simulating correlated ordinal data the `simstudy` package <https://kgoldfeld.github.io/simstudy/articles/ordinal.html> proposed a very comprehensive set of data simulation function also for ordinal data.

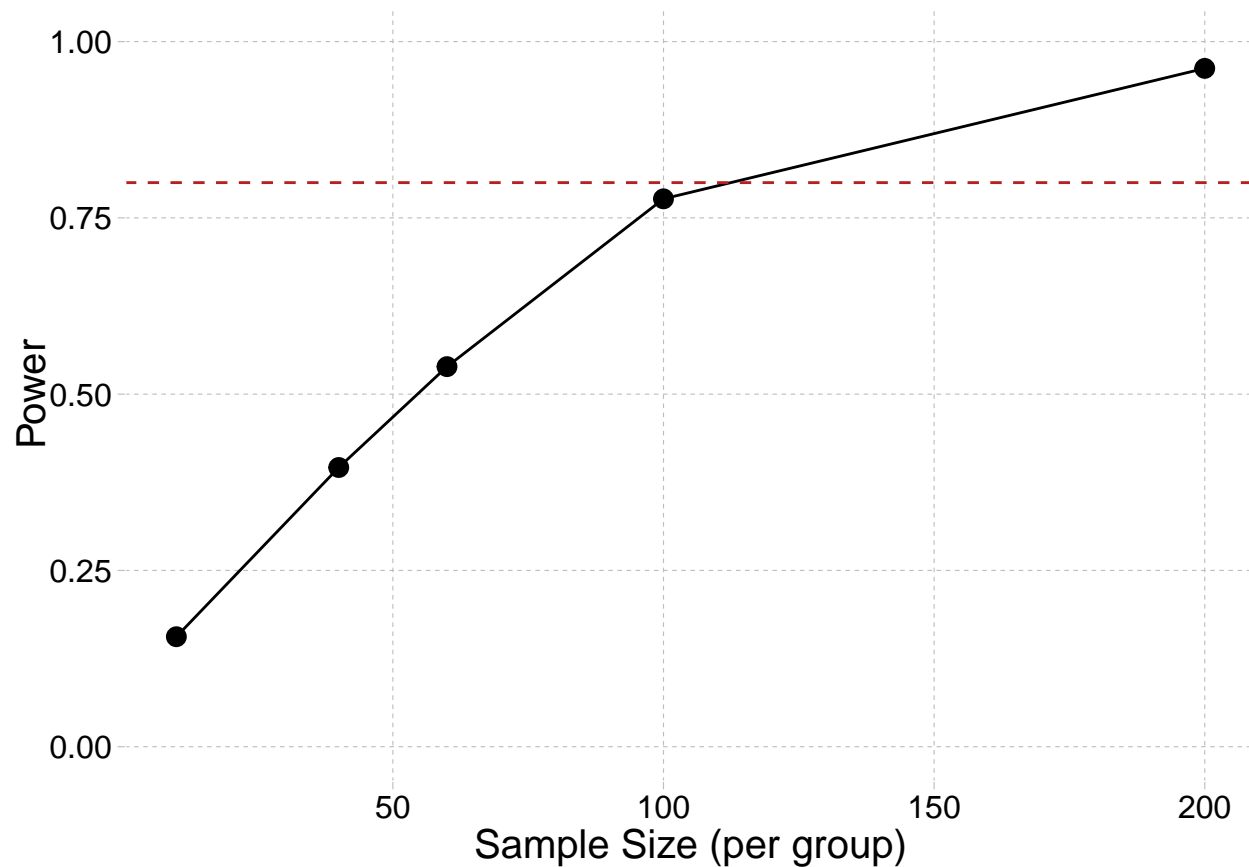


Figure 16

References

- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2, 77–101.
<https://doi.org/10.1177/2515245918823199>
- Christensen, R. H. B. (2019). *Ordinal—regression models for ordinal data*.
- Cliff, N. (2016). *Ordinal methods for behavioral data analysis*. London, England: Psychology Press.
- Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine*, 14, 1191–1203.
<https://doi.org/10.1002/sim.4780141105>
- DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological*

- Methods*, 3, 186–205. <https://doi.org/10.1037/1082-989X.3.2.186>
- DeCarlo, L. T. (2010). On the statistical and theoretical basis of signal detection theory and extensions: Unequal variance, random coefficient, and mixture models. *Journal of Mathematical Psychology*, 54, 304–313. <https://doi.org/10.1016/j.jmp.2010.01.001>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch’s *t*-test instead of student’s *t*-test. *International Review of Social Psychology*, 30, 92. <https://doi.org/10.5334/irsp.82>
- Faraggi, D., Izikson, P., & Reiser, B. (2003). Confidence intervals for the 50 per cent response dose. *Statistics in Medicine*, 22, 1977–1988. <https://doi.org/10.1002/sim.1368>
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press. <https://doi.org/10.1017/9781139161879>
- Green, D. M., Swets, J. A., & Others. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. xi, 455–xi, 455). Oxford, England: Wiley New York.
- Harrell, F. E. (2015). *Regression modeling strategies*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-19425-7>
- Kemp, S., & Grace, R. C. (2021). Using ordinal scales in psychology. *Methods in Psychology (Online)*, 5, 100054. <https://doi.org/10.1016/j.metip.2021.100054>
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in r*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-4475-6>
- Kruschke, J. K. (2015). *Doing bayesian data analysis*. <https://doi.org/10.1016/c2012-0-00477-2>
- Kurz, A. S. (2023). *Doing bayesian data analysis in brms and the tidyverse* (Version 1.1.0). Retrieved from <https://bookdown.org/content/3686/>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*,

- 22, 55. Retrieved from <https://psycnet.apa.org/record/1933-01885-001>
- Liu, A., He, H., Tu, X. M., & Tang, W. (2023). On testing proportional odds assumptions for proportional odds models. *General Psychiatry*, 36, e101048.
<https://doi.org/10.1136/gpsych-2023-101048>
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, 42, 109–127. <https://doi.org/10.1111/j.2517-6161.1980.tb01109.x>
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 54, 507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*, 5, 589965.
<https://doi.org/10.3389/educ.2020.589965>
- Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448–467.
<https://doi.org/10.1037/1082-989X.8.4.448>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc.*, 31, 137–149. <https://doi.org/10.3758/bf03207704>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science (New York, N.Y.)*, 103, 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Tutz, Gerhard. (1990). Sequential item response models with an ordered response. *The British Journal of Mathematical and Statistical Psychology*, 43, 39–55.
<https://doi.org/10.1111/j.2044-8317.1990.tb00925.x>

Tutz, Gerhard. (2022). Ordinal regression: A review and a taxonomy of models. *Wiley Interdisciplinary Reviews. Computational Statistics*, 14.

<https://doi.org/10.1002/wics.1545>

Tutz, G., & Berger, M. (2017). Separating location and dispersion in ordinal regression models. *Econometrics and Statistics*, 2, 131–148.

<https://doi.org/10.1016/j.ecosta.2016.10.002>

Tutz, Gerhard, & Berger, M. (2020). Non proportional odds models are widely dispensable – sparser modeling based on parametric and additive location-shift approaches. *arXiv [Stat.ME]*. Retrieved from <http://arxiv.org/abs/2006.03914>