



Regression for ordinal outcomes

E-mail distributed on 08-03-2024

Dear all,

In social science, we often work with outcome variables that are not “properly” continuous, for example variables that **(a)** take on just a few discrete values (e.g., 1 to 5 agreement), **(b)** have limits (e.g., 0 to 100 on a continuous slider), or **(c)** have unequally spaced levels (e.g., 1=never, 2=rarely, 3=sometimes, 4=often, and 5=all the time). Sometimes all of these issues are present simultaneously. Such data are generally called ordinal, and the question arises whether it is appropriate to use them as outcome variables in regression and, if not, what the alternative is.

Firstly, one should always keep in mind that normal linear regression is [relatively robust against non-normal data](#). That is, when the data are not severely non-linear, multi-modal, or affected by outliers, the linear model will yield valid parameter estimates and inferential tests. Moreover, because of its mathematical simplicity and ease of interpretation, I believe it should often be preferred over complex but more “correct” alternatives. I have long argued against distributional fetishism, that is, unless the research objective demands it, modelling the distribution of an outcome *exactly* is typically not necessary. The linear normal approximation will be quite good.

However, there are scenarios where one may prefer to model an ordinal outcome without normal regression. Before discussing the options, I briefly caution what *not* to use.

Not Poisson regression!

It is sometimes believed that Poisson regression should be applied to ordinal outcomes. However, this should be avoided! Poisson regression treats outcome values fundamentally different than does normal regression. For example, when one has an observation of “5”, then normal regression will treat it as a single realization of that value, with the sample size equal to 1. Poisson regression, on the other hand, will treat it as five realizations of an event, with the sample size equal to 5.

Poisson regression is appropriate for so-called “count data”, such as arise in frequency tables. These counts should be (independent) realizations of a certain event, within a certain window (e.g., number of customers per hour, number of eggs per nest). The goal of Poisson regression is to model the rate at which this event occurs, as a function of one or more predictor variables. This would clearly not be the goal for most psychological questionnaire scales (e.g., Likert), where the interest is in modelling changes in the magnitude of the outcome. Simply because a variable is discrete or integer-valued does not make it a “count” in the Poisson sense.

Poisson regression is best applied for the analysis of multi-dimensional frequency tables, where it is a generalization of chi-square analysis, known as log-linear analysis. For more information on this application, I refer to my [workshop on logistic regression](#), specifically Part 1, slides 41–64.

Ordinal regression

One drawback of modelling ordinal outcomes with normal regression is that the model may produce predicted values outside of the outcome's limits. If prediction or forecasting is the primary purpose of the model, then this will be undesirable behavior. Likewise, when there is unequal spacing between the levels of the outcome variable, a linear relationship with the predictor variables may not be plausible. In these scenarios, one could run a proper ordinal regression.

While these models are obscure to most researchers, there is a surprising diversity of choice and R packages to fit them. The three main types of ordinal models are **(a)** cumulative link models, **(b)** adjacent categories models, and **(c)** continuation/stop ratio models (Agresti, 2002). The first type is the more common one, and known as **proportional odds logistic regression (POLR)** when the link function is a logit-transformation. This model is essentially a logistic regression, with separate intercepts for shifting between successive outcome classes. Versions of it have been implemented in packages `rms`, `VGAM`, `MASS`, and `ordinal`. The last three have preference since they are supported by package `emmeans`, which can calculate condition means, and their pairwise contrasts.

For example, let us consider the following toy data, where researchers are interested to find out whether the number of days since a traumatic event predicts an ordered stage of grief:

```
library(car) ; library(MASS) ; library(VGAM)
library(ordinal) ; library(emmeans) ; library(visreg)

set.seed(666)
Order <- sample(1:5, size=120, replace=TRUE)
Days_since_event <- abs(round(Order+rnorm(120, 0, 2))) * 3
Stages <- factor(Order, levels=1:5,
  labels=c("Denial", "Anger", "Bargaining", "Depression", "Acceptance"))
Gender <- sample(c("M", "F"), size=120, replace=TRUE)
Grief <- data.frame(Stages, Days_since_event, Order, Gender)
```

Normally this outcome would be purely qualitative, and submitted to a multinomial logistic regression. However, there is a specific hypothesis in this case that the stages of grief are *ordered*. Of course, it would be tempting to use a continuous regression to simply predict the order of the stages (1–5), e.g.:

```
ols <- lm(Order~Days_since_event+Gender, data=Grief)
summary(ols)
Anova(ols, type=2)
visreg(ols, xvar="Days_since_event")
```

```

> Coefficients:
>
> Estimate Std. Error t value Pr(>|t|)
> (Intercept) 1.80053 0.21332 8.440 9.78e-14 ***
> Days_since_event 0.12386 0.01633 7.586 8.67e-12 ***
> GenderM -0.13938 0.19848 -0.702 0.484
> ---
> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> Residual standard error: 1.081 on 117 degrees of freedom
> Multiple R-squared: 0.3297, Adjusted R-squared: 0.3182
> F-statistic: 28.77 on 2 and 117 DF, p-value: 6.877e-11

```

Highly significant Days effect, supporting the idea that passing time predicts later stages of grief. However, this model is awkward to generate predicted values from, since these will not be integer values and may exceed the bounds of the outcome. Especially in this case this is highly undesirable, since, e.g., stages 0 or 6 simply do not exist. Rounding the prediction is not ideal either. Instead, we can consider a proportional odds logistic regression with the `polr()` function from the `MASS` package:

```

or <- polr(Stages~Days_since_event+Gender, data=Grief)
summary(or)
Anova(or, type=2)

```

```

> Coefficients:
>
> Value Std. Error t value
> Days_since_event 0.2079 0.03299 6.3035
> GenderM -0.2487 0.33615 -0.7398
>
> Intercepts:
>
> Value Std. Error t value
> Denial|Anger -0.0304 0.3824 -0.0794
> Anger|Bargaining 1.4033 0.3937 3.5641
> Bargaining|Depression 2.6078 0.4385 5.9475
> Depression|Acceptance 4.3241 0.5329 8.1144
>
> Residual Deviance: 333.9665
> AIC: 345.9665

```

The most striking feature of this output is the model's multiple intercepts, which control the shift between the ordinal outcome levels. The Days and Gender effect have just one coefficient, which means that, regardless between which pair of outcome levels one transitions, the odds multiply at the same rate as a function of Days, hence the name *proportional odds* logistic regression. Unfortunately, the output has no *p*-values, although we can get an ANOVA breakdown using the `Anova()` function of the `car` package. The `visreg` library allows visualization of the effect:

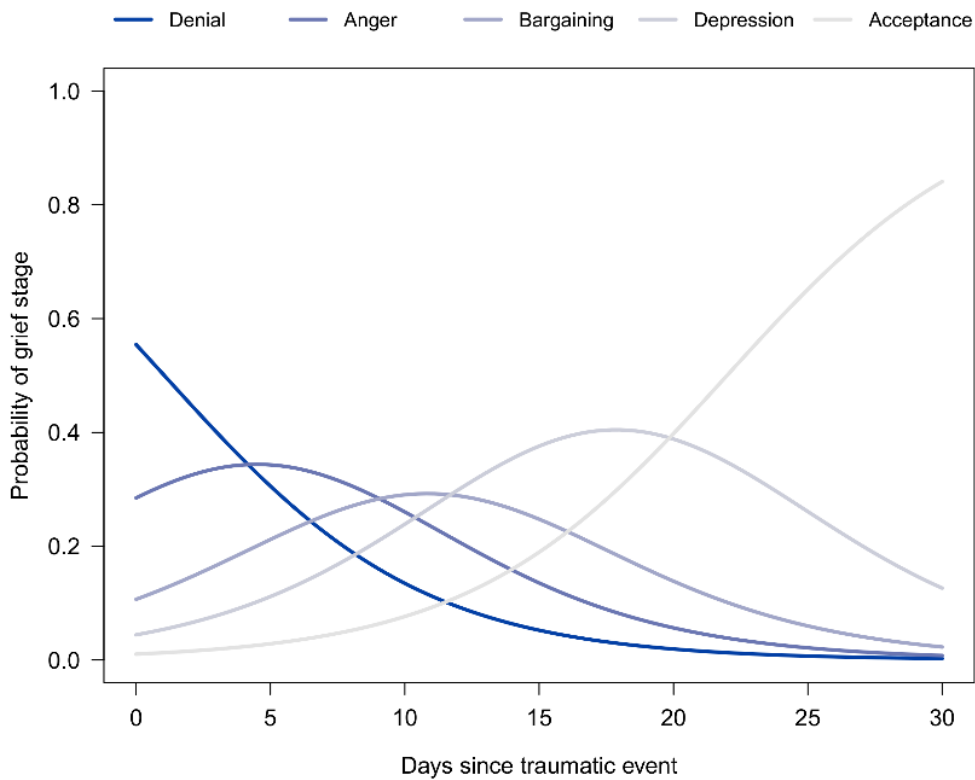


Figure 1. Predicted POLR probabilities for ordinal stages of grief as a function of days since the traumatic event.

Which illustrates nicely how stages of grief shift as the time from the traumatic event passes. We can get such predictions also with clever use of the `emmeans` functions:

```
emmeans(or, ~Stages | Days_since_event, at=list(Days_since_event=c(0,15,30)),
        mode="prob")
```

```
> Days_since_event = 0:
> Stages      prob      SE  df  asymp.LCL  asymp.UCL
> Denial      0.52339 0.08879 Inf  0.349369  0.6974
> Anger       0.29754 0.06016 Inf  0.179633  0.4154
> Bargaining  0.11758 0.03611 Inf  0.046813  0.1883
> Depression  0.04985 0.01970 Inf  0.011227  0.0885
> Acceptance  0.01165 0.00601 Inf -0.000133  0.0234
>
> Days_since_event = 15:
> Stages      prob      SE  df  asymp.LCL  asymp.UCL
> Denial      0.04662 0.01621 Inf  0.014840  0.0784
> Anger       0.12328 0.03003 Inf  0.064421  0.1821
> Bargaining  0.23492 0.04285 Inf  0.150945  0.3189
> Depression  0.38519 0.05663 Inf  0.274197  0.4962
> Acceptance  0.20999 0.04884 Inf  0.114271  0.3057
>
> Days_since_event = 30:
> Stages      prob      SE  df  asymp.LCL  asymp.UCL
```

```

> Denial      0.00216 0.00170 Inf -0.001183    0.0055
> Anger       0.00683 0.00471 Inf -0.002404    0.0161
> Bargaining 0.02036 0.01264 Inf -0.004419    0.0451
> Depression 0.11445 0.06033 Inf -0.003789    0.2327
> Acceptance 0.85620 0.07645 Inf  0.706370    1.0060

```

This lists the predicted probabilities of the five stages for three specified values of Days, illustrating that denial has highest probability at day 0, and acceptance highest probability at day 30. Two peculiarities about the use of emmeans here is **(a)** the appearance of the outcome variable in the emmeans formula, and **(b)** the mode argument for controlling the type of prediction to be generated. It is strongly advised to read the [package's vignette on ordinal models](#) for more information!

The VGAM package offers an interesting extension of POLR models, by allowing both the intercepts *and* the slopes to change as a function of shifting outcome classes. Doing so in fact allows us to make a formal comparison between a model that constrains the odds to be proportional and one that does not, e.g.:

```

POLR <- vglm(Stages~Days_since_event, data=Grief,
  family=cumulative(link=logitlink,parallel=TRUE))
NPOLR <- vglm(Stages~Days_since_event, data=Grief,
  family=cumulative(link=logitlink,parallel=FALSE))
AIC(POLR)
AIC(NPOLR)

> 344.5151
> 346.9989

```

Worse AIC for unconstrained slopes, so we accept the proportional odds assumption for these data! The package ordinal, finally, offers POLR with the c1m() function (cumulative link models), as well as multilevel POLR with the c1mm() function, which allows data from repeated measures designs to be analyzed likewise.

Best,
Ben

--

Ben Meuleman, Ph.D.
Statistician

Swiss Center for Affective Sciences
University of Geneva | Campus Biotech
Chemin des Mines 9 | CH-1202 Genève
ben.meuleman@unige.ch | +41 (0)22 379 09 79