Ordinal regression

Dae-Jin Lee
dlee@bcamath.org

Basque Center for Applied Mathematics

May 14-21, 2015

Ordinal regression

Introduction

- There are occassions when the scale of the multiple category outcome is not nominal but ordinal
- An **ordinal variable**, is a variable whose value exists on an arbitrary scale where only the relative ordering between different values is significant

  e.g.: *from low to high, or more or less of something.*
- We could use the multinomial logistic regression model in previous session, but then we would not take the ordinal nature of the outcome and then the estimated odds ratios may not address the questions of interests.

Ordinal regression

Case study: Graduate school application

- A study looks at factors that influence the decision of whether to apply to graduate school. College juniors are asked if they are `unlikely`, `somewhat likely`, or `very likely` to apply to graduate school. Hence, our outcome variable has three categories.
- Data on **parental educational status**, whether the undergraduate institution is `public` or `private`, and current GPA is also collected.
- **The researchers have reason to believe that the distances between these three points are not equal**. For example, the distance between unlikely and somewhat likely may be shorter than the distance between somewhat likely and very likely.

See `Ordreg.R` file

```
> library(foreign)
>  gradschool<-read.dta(file="data/gradschool.dta")
```

Ordinal regression

Case study: Graduate school application

- The variables are

`apply` with levels `unlikely`, `somewhat likely`, or `very likely` to apply to graduate school

`pared` which is a $0/1$ variable indicating whether at least one parent has a graduate degree

`public` which is a $0/1$ variable where 1 indicates that the undergraduate institution is public and 0 private

`gpa` which is the student's grade point average

Ordinal regression

Case study: Graduate school application

- The variables are

`apply` with levels `unlikely`, `somewhat likely`, or `very likely` to apply to graduate school

`pared` which is a $0/1$ variable indicating whether at least one parent has a graduate degree

`public` which is a $0/1$ variable where 1 indicates that the undergraduate institution is public and 0 private

`gpa` which is the student's grade point average

- Given this data, we can try different models as we learnt in this course:

Ordinal regression

Case study: Graduate school application

- The variables are

apply with levels `unlikely`, `somewhat likely`, or `very likely` to apply to graduate school

pared which is a $0/1$ variable indicating whether at least one parent has a graduate degree

public which is a $0/1$ variable where 1 indicates that the undergraduate institution is public and 0 private

gpa which is the student's grade point average

- Given this data, we can try different models as we learnt in this course:
  1. Code the data to binary $0/1$ response and fit a **logistic regression model**.
  2. Analyze each category separately and fit a **multinomial logistic regression model**.

Ordinal regression

Case study: Graduate school application

- The variables are

`apply` with levels `unlikely`, `somewhat likely`, or `very likely` to apply to graduate school

`pared` which is a $0/1$ variable indicating whether at least one parent has a graduate degree

`public` which is a $0/1$ variable where 1 indicates that the undergraduate institution is public and 0 private

`gpa` which is the student's grade point average

- Given this data, we can try different models as we learnt in this course:
  1. Code the data to binary $0/1$ response and fit a **logistic regression model**.
  2. Analyze each category separately and fit a **multinomial logistic regression model**.
- Both models can be used, but they will be not efficient, as we are not using all the information of the outcomes (**rank ordering**).

Ordinal regression

Proportional odds model

- Essentially, ordinal regression is another extension of logistic regression which is based on a model called the **proportional odds model**.

Ordinal regression

Proportional odds model

- Essentially, ordinal regression is another extension of logistic regression which is based on a model called the **proportional odds model**.
- What this model does is actually turn the ordinal scale into a number of **binary cut-off points**, the number of cut-off points always being one less than the number of categories.

Ordinal regression

Proportional odds model

- Essentially, ordinal regression is another extension of logistic regression which is based on a model called the **proportional odds model**.
- What this model does is actually turn the ordinal scale into a number of **binary cut-off points**, the number of cut-off points always being one less than the number of categories.
- The **cut-off points** can be viewed as **thresholds** we need to cross to get to the next (higher) category from the category below, which means, that in a four point ordinal scale variable (coded `4 = agree strongly`, `3=agree`, `2=disagree`, `1=disagree strongly`), for example, we'd have three thresholds, one to go from category 1 to 2, one to go from 2 to 3 and one to go from 3 to 4. Then we estimate the probability of observing a particular score or less.

Ordinal regression

Proportional odds model

- Essentially, ordinal regression is another extension of logistic regression which is based on a model called the **proportional odds model**.
- What this model does is actually turn the ordinal scale into a number of **binary cut-off points**, the number of cut-off points always being one less than the number of categories.
- The **cut-off points** can be viewed as **thresholds** we need to cross to get to the next (higher) category from the category below, which means, that in a four point ordinal scale variable (coded `4 = agree strongly`, `3=agree`, `2=disagree`, `1=disagree strongly`), for example, we'd have three thresholds, one to go from category 1 to 2, one to go from 2 to 3 and one to go from 3 to 4. Then we estimate the probability of observing a particular score or less.
- We estimate separate binary regression models for each of:
  - Pr(score of 1)/Pr(score greater than 1)
  - Pr(score of 1 or 2)/Pr(score greater than 2)
  - Pr(score of 1, 2 or 3)/Pr(score greater than 3)

Ordinal regression

Proportional odds model

- The **proportional odds model** is a class of GLM used for modeling the dependence of an ordinal response on discrete or continuous covariates X.
- Let Y denote the response category in the range $1, ..., K,$ with $K \geq 2$, and let

$$\gamma_k = \mathrm{Pr}(Y \leq k|X), \quad \text{for } k = 1, ..., K-1$$

  be the **cumulative response probability** given the value of the covariate X and k the **cut-off points**.
- The most general form of a proportional odds model for the k^{th} **cumulative response probability** is:

$$\mathrm{logit}(\gamma_k) = \alpha_k - \beta x$$

  where the intercept α_k depend on the category k and the slopes β are equal for all the categories.

Ordinal regression

Proportional odds model

- The model compares the probability of an outcome $\leq k$ to an outcome $> k$.
- Then

$$\text{logit}(\gamma_1) \quad = \quad \log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 - \beta_1 x_1 + \ldots - \beta_m x_m$$

Ordinal regression

Proportional odds model

- The model compares the probability of an outcome $\leq k$ to an outcome $> k$.
- Then

$$\begin{aligned}\text{logit}(\gamma_1) &= \log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 - \beta_1 x_1 + \ldots - \beta_m x_m \\ \text{logit}(\gamma_2) &= \log\left(\frac{p_1+p_2}{1-p_1-p_2}\right) = \alpha_2 - \beta_1 x_1 + \ldots - \beta_m x_m\end{aligned}$$

Ordinal regression

Proportional odds model

- The model compares the probability of an outcome $\leq k$ to an outcome $> k$.
- Then

$$
\begin{aligned}
\text{logit}(\gamma_1) &= \log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 - \beta_1 x_1 + \ldots - \beta_m x_m \\
\text{logit}(\gamma_2) &= \log\left(\frac{p_1+p_2}{1-p_1-p_2}\right) = \alpha_2 - \beta_1 x_1 + \ldots - \beta_m x_m \\
\text{logit}(\gamma_k) &= \log\left(\frac{p_1+p_2+\ldots+p_k}{1-p_1-p_2-\ldots-p_k}\right) = \alpha_k - \beta_1 x_1 + \ldots - \beta_m x_m
\end{aligned}
$$

Ordinal regression

Proportional odds model

- The model compares the probability of an outcome $\leq k$ to an outcome $> k$.
- Then

$$
\begin{aligned}
\text{logit}(\gamma_1) &= \log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 - \beta_1 x_1 + \ldots - \beta_m x_m \\
\text{logit}(\gamma_2) &= \log\left(\frac{p_1+p_2}{1-p_1-p_2}\right) = \alpha_2 - \beta_1 x_1 + \ldots - \beta_m x_m \\
\text{logit}(\gamma_k) &= \log\left(\frac{p_1+p_2+\ldots+p_k}{1-p_1-p_2-\ldots-p_k}\right) = \alpha_k - \beta_1 x_1 + \ldots - \beta_m x_m \\
1 &= p_1 + p_2 + \ldots + p_k + p_{k+1}
\end{aligned}
$$

where α's are different.

Proportional odds model

Why $-\beta_m$ instead of $+\beta_m$?

- Analogously to logistic regressiom where the outcome W takes values $0/1$:

$$p = \text{Pr}(W = 1|x) \quad \text{then} \quad \text{logit}(p) = \beta_1 x$$

- If we collapse the categories of the ordinal outcome Y (e.g.: 1,2, and 3) into binary:

$$\underbrace{1, 2}_{0} \qquad \underbrace{3}_{1}$$

Proportional odds model

Why $-\beta_m$ instead of $+\beta_m$?

- Analogously to logistic regressiom where the outcome W takes values $0/1$:

$$p = \Pr(W = 1|x) \quad \text{then} \quad \text{logit}(p) = \beta_1 x$$

- If we collapse the categories of the ordinal outcome Y (e.g.: 1,2, and 3) into binary:

$$\underbrace{1,2}_{0} \qquad \underbrace{3}_{1}$$

- Then

$\Pr(Y \leq 2|x) = \Pr(W = 0|x) = 1-p$

Proportional odds model

Why $-\beta_m$ instead of $+\beta_m$?

- Analogously to logistic regressiom where the outcome W takes values $0/1$:

$$p = \mathrm{Pr}(W = 1|x) \quad \text{then} \quad \mathrm{logit}(p) = \beta_1 x$$

- If we collapse the categories of the ordinal outcome Y (e.g.: 1,2, and 3) into binary:

$$\underbrace{1,2}_{0} \qquad \underbrace{3}_{1}$$

- Then

$$\mathrm{Pr}(Y \leq 2|x) = \mathrm{Pr}(W = 0|x) = 1-p \Rightarrow \mathrm{logit}(\mathrm{Pr}(Y \leq 2|x)) = \log\left(\frac{1-p}{p}\right) = -\beta_1 x$$

Proportional odds model

Why $-\beta_m$ instead of $+\beta_m$?

- Analogously to logistic regressiom where the outcome W takes values $0/1$:

$$p = \text{Pr}(W = 1|x) \quad \text{then} \quad \text{logit}(p) = \beta_1 x$$

- If we collapse the categories of the ordinal outcome Y (e.g.: 1,2, and 3) into binary:

$$\underbrace{1,2}_{0} \qquad \underbrace{3}_{1}$$

- Then

$$\text{Pr}(Y \leq 2|x) = \text{Pr}(W = 0|x) = 1-p \Rightarrow \text{logit}(\text{Pr}(Y \leq 2|x)) = \log\left(\frac{1-p}{p}\right) = -\beta_1 x$$

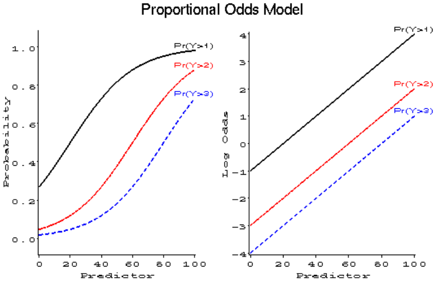
- Then, **each β is interpreted as the change in the log-odds from the highest category to a lower category, associated to an increment of a unit of X, with all the rest of covariates remain constant.**
- This is true for any of the rest of categories.

Proportional odds model

- The model is called **proportional odds**, because the odds for any k is proportional to

$$e^{-\beta_1 x_1 - \beta_2 x_2 - \ldots - \beta_m x_m}$$

e.g.:



- The model assumes that the regression functions for different response categories are parallel on the logit scale.

Proportional odds model

Interpretation

- In this model, intercept α_k is the log-odds of falling into or below category k when $x_1 = x_2 = ... = 0$.
- A single parameter β_j describes the effect of x_j on Y such that β_j is the increase in log-odds of **falling into or below** any category associated with a one-unit increase in x_j, holding all the other X-variables constant; compare this to the baseline logit model where there are $K-1$ parameters for a single explanatory variable. Therefore, a positive slope indicates a tendency for the response level to decrease as the variable decreases.
- But when we fit the model, we reverse the order of the response categories so that a positive slope correspond to increases in the response.
- **Constant slopes** β_j: The effect of x_j, is the same for all $K-1$ ways to collapse Y into dichotomous outcomes.
- The odds-ratio is proportional to the difference between x_1 and x_2 where β is the constant of proportionality: $\exp[\beta(x_1 - x_2)]$, and thus the name "proportional odds model".

Proportional odds model

as a continuous latent variable

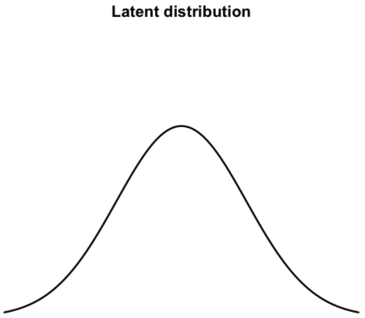
- One reason for the proportional-odds cumulative-logit model's popularity is its connection to the idea of a **continuous latent response**.
- Suppose that the categorical outcome is actually a categorized version of an unobservable (latent) continuous variable, Z, that we discretized as:

$$c_{j-1} < Z \leq c_j \Rightarrow Y = j \qquad j = 1, ..., k+1$$

Proportional odds model

as a continuous latent variable

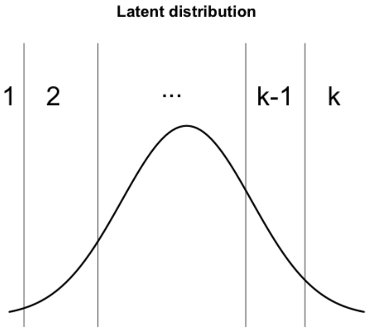
- One reason for the proportional-odds cumulative-logit model's popularity is its connection to the idea of a **continuous latent response**.
- Suppose that the categorical outcome is actually a categorized version of an unobservable (latent) continuous variable, Z, that we discretized as:



Proportional odds model

as a continuous latent variable

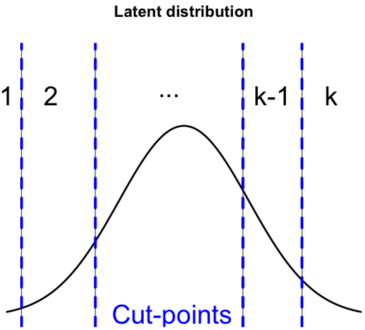
- One reason for the proportional-odds cumulative-logit model's popularity is its connection to the idea of a **continuous latent response**.
- Suppose that the categorical outcome is actually a categorized version of an unobservable (latent) continuous variable, Z, that we discretized as:



Proportional odds model

as a continuous latent variable

- One reason for the proportional-odds cumulative-logit model's popularity is its connection to the idea of a **continuous latent response**.
- Suppose that the categorical outcome is actually a categorized version of an unobservable (latent) continuous variable, Z, that we discretized as:



Proportional odds model

as a continuous latent variable (cont.)

- For example, it is reasonable to think that a 5-point scale (`1 = strongly disagree`, `2 = agree`, `3 = neutral`, `4 = agree`, `5 = strongly agree`) is a coarsened version of a continuous variable Z indicating degree of approval.

Proportional odds model

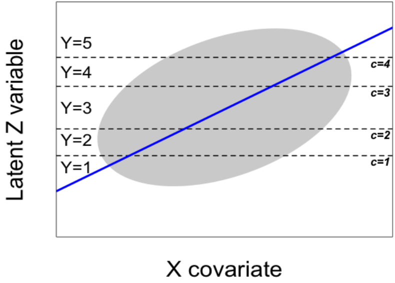
as a continuous latent variable (cont.)

- For example, it is reasonable to think that a 5-point scale (`1 = strongly disagree`, `2 = agree`, `3 = neutral`, `4 = agree`, `5 = strongly agree`) is a coarsened version of a continuous variable Z indicating degree of approval.
- The continuous scale is divided into five regions by four cut-points c_1, c_2, c_3, c_4 which are determined by nature (not by the investigator). If $Z \leq c_1$ we observe $Y = 1$; if $c_1 < Z \leq c_2$ we observe $Y = 2$; and so on.

Proportional odds model

as a continuous latent variable (cont.)

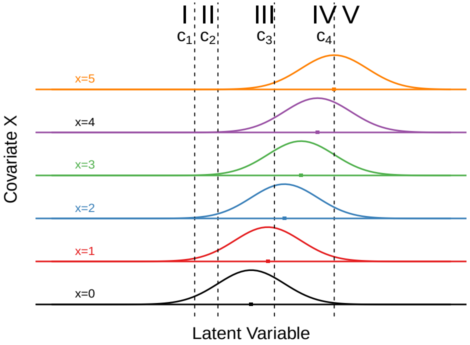
- For example, it is reasonable to think that a 5-point scale (`1 = strongly disagree`, `2 = agree`, `3 = neutral`, `4 = agree`, `5 = strongly agree`) is a coarsened version of a continuous variable Z indicating degree of approval.
- The continuous scale is divided into five regions by four cut-points c_1, c_2, c_3, c_4 which are determined by nature (not by the investigator). If $Z \leq c_1$ we observe $Y = 1$; if $c_1 < Z \leq c_2$ we observe $Y = 2$; and so on.
- Here is the connection: Suppose that Z is related to the X's through a linear regression.



- The regression of Z on the X covariates has the form

$$Z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon,$$

where ϵ is a random error from a logistic distribution with mean zero and constant variance, then the coarsened version Y will be related to the X's by a proportional-odds cumulative logit model (with homocedastic errors).



- How the distribution of the latent variable Z changes with x.
- Recorded categories are denoted by **I,II,III,IV and V** categories attached to the 5 contiguous Z-intervals.

- ϵ has a distribution function F such that:

$$\mathsf{Pr}(Y \leq j) = \mathsf{Pr}(Z \leq \alpha_j) = F(\alpha_j - \beta_1 x_1 - \beta_2 x_2 - ... - \beta_p x_p)$$

  and

$$F^{-1}(p_i + \ldots + p_j) = \alpha_j - \beta_1 x_1 - \beta_2 x_2 - ... - \beta_k x_p$$

  if F^{-1} is the **logit**, the have the proportional odds model.

- ϵ has a distribution function F such that:

$$\Pr(Y \leq j) = \Pr(Z \leq \alpha_j) = F(\alpha_j - \beta_1 x_1 - \beta_2 x_2 - ... - \beta_p x_p)$$

  and

$$F^{-1}(p_i + \ldots + p_j) = \alpha_j - \beta_1 x_1 - \beta_2 x_2 - ... - \beta_k x_p$$

  if F^{-1} is the **logit**, the have the proportional odds model.

- The odds of the event $Y \leq j$ satisfies

$$\text{odds}(Y \leq j|x) = \exp(\alpha_j - \beta x)$$

  Consequently, the Odds ratio of the event $Y \leq j$ for x_1 and x_0 is

$$\frac{\text{Odds}(Y \leq j|X = x_1)}{\text{Odds}(Y > j|X = x_1)} = \exp\left(\beta(x_1 - x_0)\right)$$

  which is a constant independent of k.

- If x_0 is the baseline value of the covariate, it follows that $\exp(\alpha_j)$ is the baseline odds for the event $Y \leq j$.

Proportional odds model

function polr in library(MASS)

- The command name comes from proportional odds logistic regression.

See Ordreg.R

```
> library(MASS)
> ord1 <- polr(apply~pared,data=gradschool)
> summary(ord1)

Call:
polr(formula = apply ~ pared, data = gradschool)

Coefficients:
       Value Std. Error t value
pared1 1.127     0.2634    4.28

Intercepts:
                            Value   Std. Error t value
unlikely|somewhat likely     0.3768  0.1103      3.4152
somewhat likely|very likely  2.4519  0.1826     13.4302

Residual Deviance: 722.7903
AIC: 728.7903
```

Proportional odds model

function `polr` in `library(MASS)`

- The output show two intercepts (cut-off points), they indicate where the latent variable is cut to make the three groups that we observe in our data.

```
Coefficients:
      Value Std. Error t value
pared 1.127     0.2634    4.28

Intercepts:
                            Value   Std. Error t value
unlikely|somewhat likely     0.3768  0.1103     3.4152
somewhat likely|very likely  2.4519  0.1826    13.4302
```

- The coefficient for `pared` is positive, and indicates that having at least a parent having a graduate degree increases the chances to apply to grad school.
- We can compute the p-values manually and obtain CI's (See `GLM-BCAM-Ordreg.R` script)

Proportional odds model

function `polr` in `library(MASS)`

- Let us consider the other covariates
- **Which covariates are significative?**

Proportional odds model

function `polr` in `library(MASS)`

- Let us consider the other covariates
- **Which covariates are significative?** $\Rightarrow$ **LRT**

```
> ord2 <- polr(apply~public,data=gradschool)
> summary(ord2)
> ord3 <- polr(apply~gpa,data=gradschool)
> summary(ord3)
> ord0 <- polr(apply~1,data=gradschool)
> summary(ord0)
> anova(ord0,ord1)
> anova(ord0,ord2)
> anova(ord0,ord3)
```

Proportional odds model

function polr in library(MASS)

- Now we include only pared and gpa

```
> ord4 <- polr(apply~pared+gpa,data=gradschool)
> summary(ord4)
> pvals(ord4)
```

Proportional odds model

Propotional odds/parallel regression assumption

- One of the assumptions underlying ordinal logistic (and ordinal probit) regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption.

Proportional odds model

Propotional odds/parallel regression assumption

- One of the assumptions underlying ordinal logistic (and ordinal probit) regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption.

```
> library(VGAM)
> apply2=ordered(gradschool$apply)
> m0 <- vglm(apply2~pared+gpa,family=cumulative(parallel=TRUE),data=gradschool)
> m1 <- vglm(apply2~pared+gpa,family=cumulative(parallel=FALSE),data=gradschool)
> test.po <- 2*logLik(m1)-2*logLik(m0)
> df.po <- length(coef(m1))-length(coef(m0))
> test.po
> df.po
> 1-pchisq(test.po,df=df.po) # 0.6584501 (Do not reject the null hypothesis)
```

$*$ setting `parallel = TRUE` will fit a proportional odds model

Proportional odds model

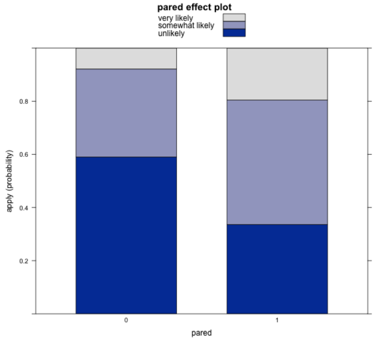
Interpretation of the odds

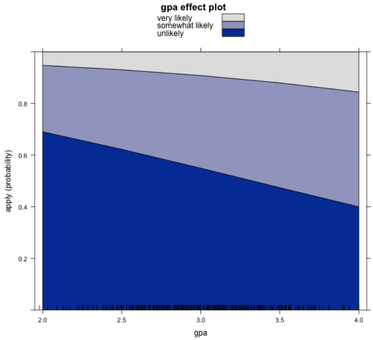
- For pared, we would say that for a one unit increase in parental education, i.e., going from 0 (Low) to 1 (High), the odds of ''very likely'' applying versus ''somewhat likely'' or ''unlikely'' applying combined are $2{,}85$ greater, given that all of the other variables in the model are held constant.
- Likewise, the odds ''very likely'' or ''somewhat likely'' applying versus ''unlikely'' applying is $2{,}85$ times greater, given that all of the other variables in the model are kept constant.
- For gpa (and other continuous variables), the interpretation is that when a student's gpa moves 1 unit, the odds of moving from ''unlikely'' applying to ''somewhat likely'' or ''very likley'' applying (or from the lower and middle categories to the high category) are multiplied by $1{,}85$.

```
exp(cbind(OR=coef(ord4),exp(confint(ord4))))
            OR     2.5 %   97.5 %
pared 2.845412 1.693056 4.806454
gpa   1.829873 1.115180 3.022320
```

- Plot of effects with `library(effects)`

```
> library(effects)
> plot(effect("pared",ord4),style="stacked")
> plot(effect("gpa",ord4),style="stacked")
```





Proportional odds model

Summary

- The steps in model building for an ordinal logistic model are the same as described for logistic regression.
- Unfortunately, the full options to check the model are not available yet in software packages.
- Diagnostic statistics and goodness-of-fit tests have not been extended for use with ordinal models.
- Thus one has to use the separate logistic regression approach. A disadvantage of this approach is that one is really not checking the actual fitted model, only an approximation to it. However, this method may help to identify influential subjects, or non-linear relationships (beyond this course).