Multinomial Regression

Dae-Jin Lee
dlee@bcamath.org

Basque Center for Applied Mathematics

http://idaejin.github.io/bcam-courses/

Introduction

Case study: mammography experience study

Multinomial regression

Introduction

- In the previous lecture we have looked at the logistic regression case when the dependent variable is **binary**, i.e. it has *only two categories*.
- Now we consider the case where we have nominal variables with more than two categories.
- We can extend the logistic regression model to do this. We call this extended method **multinomial logistic regression**.
- The basic principle of multinomial logistic regression is similar to that for (binomial) logistic regression, in that it is based on the *probability of membership of each category of the dependent variable*.
- That is why it can be viewed as a **classification method** that generalizes logistic regression to **multi-class problems**.
- **It is also known as:** Multiclass Log Reg, Multinomial Logit, or Maximum Entropy Classifier or Multinomial Choice Model.

Multinomial regression

Introduction

- We assume that the categories of the outcome variable Y, are coded $0, 1,$ or 2. In the **three outcome category** we need **two logit functions**.
- We have to decide which outcome categories to compare.
- The obvious extension is to use Y$=0$ as the **referent** or **baseline outcome** and form two logits comparing $Y = 1$ and $Y = 2$ to it.

Multinomial regression

Introduction

- We assume that the categories of the outcome variable Y, are coded $0, 1,$ or 2. In the **three outcome category** we need **two logit functions**.
- We have to decide which outcome categories to compare.
- The obvious extension is to use Y$=0$ as the **referent** or **baseline outcome** and form two logits comparing $Y = 1$ and $Y = 2$ to it.

- We denote the two logit functions as:

$$
\begin{aligned}
g_1(X) = \log\left(\frac{\Pr(Y = 1|X)}{\Pr(Y = 0|X)}\right) &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1k}x_k = \boldsymbol{X}\boldsymbol{\beta}_1 \\
g_2(X) = \log\left(\frac{\Pr(Y = 2|X)}{\Pr(Y = 0|X)}\right) &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1k}x_k = \boldsymbol{X}\boldsymbol{\beta}_2
\end{aligned}
$$

Multinomial regression

Introduction

- It follows that the conditional probabilities of each outcome category given the covariate vector are

$$\Pr(Y = 0|X) = \frac{1}{1 + \exp(g_1(X) + g_2(X))}$$

Multinomial regression

Introduction

- It follows that the conditional probabilities of each outcome category given the covariate vector are

$$\begin{aligned}
\mathsf{Pr}(Y=0|X) &= \frac{1}{1+\exp(g_1(X)+g_2(X))} \\
\mathsf{Pr}(Y=1|X) &= \frac{e^{g_1(X)}}{1+\exp(g_1(X)+g_2(X))}
\end{aligned}$$

Multinomial regression

Introduction

- It follows that the conditional probabilities of each outcome category given the covariate vector are

$$\begin{aligned}
\Pr(Y=0|X) &= \frac{1}{1+\exp(g_1(X)+g_2(X))} \\
\Pr(Y=1|X) &= \frac{e^{g_1(X)}}{1+\exp(g_1(X)+g_2(X))} \\
\Pr(Y=2|X) &= \frac{e^{g_2(X)}}{1+\exp(g_1(X)+g_2(X))}
\end{aligned}$$

Multinomial regression

Introduction

- It follows that the conditional probabilities of each outcome category given the covariate vector are

$$\begin{aligned}
\text{Pr}(Y=0|X) &= \frac{1}{1+\text{exp}(g_1(X)+g_2(X))}\\
\text{Pr}(Y=1|X) &= \frac{e^{g_1(X)}}{1+\text{exp}(g_1(X)+g_2(X))}\\
\text{Pr}(Y=2|X) &= \frac{e^{g_2(X)}}{1+\text{exp}(g_1(X)+g_2(X))}
\end{aligned}$$

- A general expression for K categories is:

$$\text{Pr}(Y=k|X) = \frac{\text{exp}(g_k(X))}{\sum_{k=0}^{K}\text{exp}(g_k(X))},$$

where $\text{Pr}(Y=k|X)$ is the probability of belonging to group k, and $\beta_0 = 0$ and $g_0(X) = 0$.

Multinomial regression

Case study

- Data from a study undertaken to assess factors associated with women's knowledge, attitude and behaviour towards the benefits of mammography (Zapka et al. 1991).
- **Source:** Hosmer and Lemeshow (2000). *Applied Logistic Regression, 2nd edition*. John Wiley Sons Inc.,New York. Section 8.1.2, page 264.

See `Multinom.R` script

```
> mamexp=read.table("data/mammexp.txt",header=TRUE)
> names(mamexp)

[1] "me"    "symp" "pb"    "hist" "bse"   "dect"
```

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

  me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

  me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

  symp Agreement with the statement: *'You do not need a mamogram unless you develop symptoms'* A factor with levels `1=Strongly Agree`, `2=Agree`, `3=Disagree` and `4=Strongly Disagree`

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

symp Agreement with the statement: *'You do not need a mamogram unless you develop symptoms'* A factor with levels `1=Strongly Agree`, `2=Agree`, `3=Disagree` and `4=Strongly Disagree`

pb Perceived benefit of mammography, the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

symp Agreement with the statement: *'You do not need a mamogram unless you develop symptoms'* A factor with levels `1=Strongly Agree`, `2=Agree`, `3=Disagree` and `4=Strongly Disagree`

pb Perceived benefit of mammography, the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

hist Mother or Sister with a history of breast cancer; a factor with levels `0=No` and `1=Yes`.

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

symp Agreement with the statement: *'You do not need a mamogram unless you develop symptoms'* A factor with levels `1=Strongly Agree`, `2=Agree`, `3=Disagree` and `4=Strongly Disagree`

pb Perceived benefit of mammography, the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

hist Mother or Sister with a history of breast cancer; a factor with levels `0=No` and `1=Yes`.

bse Answers to the question: *'Has anyone taught you how to examine your own breasts?'* A factor with levels `0=No` and `1=Yes`.

Multinomial regression

Mammography Experience Study

- A data frame with 412 observations on the following 6 variables.

me Mammograph experience, a factor with levels `0=Never`, `1=Within a Year` and `2=Over a Year`.

symp Agreement with the statement: *'You do not need a mamogram unless you develop symptoms'* A factor with levels `1=Strongly Agree`, `2=Agree`, `3=Disagree` and `4=Strongly Disagree`

pb Perceived benefit of mammography, the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

hist Mother or Sister with a history of breast cancer; a factor with levels `0=No` and `1=Yes`.

bse Answers to the question: *'Has anyone taught you how to examine your own breasts?'* A factor with levels `0=No` and `1=Yes`.

dect Answers to the question: *'How likely is it that a mammogram could find a new case of breast cancer?'* A factor with levels `0=Not likely`, `1=Somewhat likely`, `2=Very likely`.

Multinomial regression

Different alternatives in R

- There are different libraries in R with functions to fit a multinomial regression model.

Multinomial regression

Different alternatives in R

- There are different libraries in R with functions to fit a multinomial regression model.

```
> library(VGAM)
> library(nnet)
> library(mlogit)
```

- `vglm()` in `VGAM`: estimation is based on likelihood-inference
- `multinom()` in `nnet`: estimation is based on neural networks
- `mlogit()` in `mlogit`: uses a specific format, so data frame has to be reshaped

Multinomial regression

Case study

- We begin by considering a model containing a single dichotomous covariate coded 0 or 1.

Multinomial regression

Case study

- We begin by considering a model containing a single dichotomous covariate coded 0 or 1.
- In this case, in the binary outcome model the estimated slope coefficient is identical to the **log OR** obtained from the 2×2 cross-clasifying table for the outcome and the covariate.
- The cross-classification of Mammography Experience (`me`) by Family History of Breast Cander (`hist`) is

```
> xtabs(~me+hist,data=mamexp)

   hist
me     0   1
  0  220  14
  1   85  19
  2   63  11
```

- `me=0` is the reference outcome value
- $\widehat{\text{OR}}_1 = \frac{19/14}{85/220} = 3{,}51$ (i.e. $\exp(\beta_1)$)
- $\widehat{\text{OR}}_2 = \frac{11/14}{63/220} = 2{,}74$ (i.e. $\exp(\beta_2)$)

Multinomial regression

Fitting in R

- vglm is a very large class of models that includes generalized linear models (GLM's) as a special case.

```
> vglm1 <- vglm(me ~ hist, family=multinomial(refLevel=1), data=mamexp)
```

Multinomial regression

Fitting in R

- vglm is a very large class of models that includes generalized linear models (GLM's) as a special case.

```
> vglm1 <- vglm(me ~ hist, family=multinomial(refLevel=1), data=mamexp)
```

- Usually we will use multinom() in library(nnet) because it includes more options (e.g. anova()) and it is very similar to glm() function. But in general, the results are quite similar.

```
                 Value Std.Error   t value      p value
Intercept.1 -0.9509794 0.1277115 -7.446312 9.598590e-14
Intercept.2 -1.2504803 0.1428926 -8.751189 2.111168e-18
hist1.1      1.2565310 0.3746633  3.353761 7.972127e-04
hist1.2      1.0093843 0.4275097  2.361079 1.822183e-02
```

Multinomial regression

multinom()

```
> multi1


Call:
multinom(formula = me ~ hist, data = mamexp)

Coefficients:
  (Intercept)    hist1
1  -0.9509794 1.256531
2  -1.2504803 1.009384

Residual Deviance: 792.3399
AIC: 800.3399


> coefficients(multi1) # We can extract the coefficients of the model
```

Multinomial regression

multinom()

- The output is similar to logistic regression
  - By default the reference level is the 1st level.
  - Now we have two groups of parameters, one for each logit (for me=0 as baseline):

$$\begin{aligned} \text{logit}(\texttt{me=1}|\texttt{hist}) &= -0{,}95 + 1{,}25 \times \texttt{hist} \\ \text{logit}(\texttt{me=2}|\texttt{hist}) &= -1{,}25 + 1{,}01 \times \texttt{hist} \end{aligned}$$

- p-values are not provided (but can be easily obtained) $\Rightarrow$ we would need **LRT**

```
> exp(coefficients(multi1))

  (Intercept)    hist1
1   0.3863624 3.513213
2   0.2863672 2.743911
```

- Note that the **OR** are the exponentials of the β's and they identical to the values obtained in the 2×2 cross-table

Multinomial regression

multinom()

- $95\,\%$ Confidence intervals for the OR are obtained as:

```
> exp(confint(multi1))
, , 1

                 2.5 %    97.5 %
(Intercept) 0.3008061 0.4962529
hist1       1.6857397 7.3218101

, , 2

                 2.5 %    97.5 %
(Intercept) 0.2164178 0.3789254
hist1       1.1870612 6.3425948
```

- We interpret each estimated **OR** and its corresponding CI's as if it came from a logistic regression model (in few seconds).

Multinomial regression

`multinom()`

- In order to test if the predictor is significant, we use the `anova()` function, and the Likelihood Ratio Test (LRT).

Multinomial regression

multinom()

- In order to test if the predictor is significant, we use the anova() function, and the Likelihood Ratio Test (LRT).

```
> multi0 <- multinom(me~1,data=mamexp)

# weights:  6 (2 variable)
initial  value 452.628263
final  value 402.599009
converged

> anova(multi0,multi1)

Likelihood ratio tests of Multinomial Models

Response: me
  Model Resid. df Resid. Dev   Test    Df LR stat.     Pr(Chi)
1     1       822   805.1980
2  hist       820   792.3399 1 vs 2     2 12.85808 0.001614001
```

- Hence the varible hist is significant

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with family history of breast cancer are $3{,}51$ times more likely to be frequent users of mammography screening than are women without a family history of breast cancer.

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with family history of breast cancer are $3{,}51$ times more likely to be frequent users of mammography screening than are women without a family history of breast cancer.
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `More than a Year ago` is $2{,}74$ times greater than the odds among women without family history (`HIST=No`).

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with family history of breast cancer are $3{,}51$ times more likely to be frequent users of mammography screening than are women without a family history of breast cancer.
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `More than a Year ago` is $2{,}74$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with a history of breast cancer are $2{,}7$ times as likely to have had a mammogram over one year ago than women without a family history of breast cancer.

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with family history of breast cancer are $3{,}51$ times more likely to be frequent users of mammography screening than are women without a family history of breast cancer.
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `More than a Year ago` is $2{,}74$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with a history of breast cancer are $2{,}7$ times as likely to have had a mammogram over one year ago than women without a family history of breast cancer.
- Thus, `hist` is a significant factor in use of mammography screening.

Multinomial regression

Interpretation of the model

- The interpretation of `hist` ("Family history of breast cancer") on the frequency of a mammography screening `me` is as follows
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `Within a Year` is $3{,}51$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with family history of breast cancer are $3{,}51$ times more likely to be frequent users of mammography screening than are women without a family history of breast cancer.
  - The odds among women with a family history of breast cancer (`HIST=Yes`) having a mammogram `More than a Year ago` is $2{,}74$ times greater than the odds among women without family history (`HIST=No`).
    $*$ In other words, women with a history of breast cancer are $2{,}7$ times as likely to have had a mammogram over one year ago than women without a family history of breast cancer.
- Thus, `hist` is a significant factor in use of mammography screening.
- **If we include instead of `HIST`, the variable `DECT` (women opinion towards a mammography)?**

Multinomial regression

polychotomous covariate

- We now consider the variable `dect` that has three levels `0=Not likely`, `1=Somewhat likely` and `2=Very likely`

Multinomial regression

polychotomous covariate

- We now consider the variable `dect` that has three levels `0=Not likely`, `1=Somewhat likely` and `2=Very likely`

Cross classification table of `me` by `dect`:

```
> xtabs(~me+dect,data=mamexp)

   dect
me    0   1   2
  0  13  77 144
  1   1  12  91
  2   4  16  54
```

- `me=0` is the reference outcome value and `dect=0` as the reference covariate values
  - $\widehat{\mathrm{OR}}_1(1,0) = \frac{12\times 13}{77\times 1} = 2{,}03$
  - $\widehat{\mathrm{OR}}_1(2,1) = \frac{91\times 13}{144\times 1} = 8{,}22$
  - $\widehat{\mathrm{OR}}_2(1,0) = \frac{16\times 13}{77\times 4} = 0{,}68$
  - $\widehat{\mathrm{OR}}_2(2,1) = \frac{54\times 13}{144\times 4} = 1{,}22$

Multinomial regression

```
> multi2 <- multinom(me~dect,data=mamexp)

> anova(multi0,multi2)

Likelihood ratio tests of Multinomial Models

Response: me
  Model Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1     1       822   805.1980
2  dect       818   778.4011 1 vs 2     4 26.79694 2.184912e-05
```

- Thus, woman's opinion on the ability of a mammogram to detect a new case of breast cancer is significantly associated to her decision to have a mammogram.

```
> coef(multi2)

  (Intercept)      dect1     dect2
1   -2.565201  0.7062589 2.1062453
2   -1.178617 -0.3926042 0.1977986

> exp(coef(multi2))

  (Intercept)     dect1    dect2
1  0.07690374 2.0263961 8.217330
2  0.30770391 0.6752959 1.218717
```

```
> exp(confint(multi2))

, , 1

                 2.5 %     97.5 %
(Intercept) 0.01005804  0.5880059
dect1       0.24245780 16.9360656
dect2       1.05675131 63.8982038

, , 2

                2.5 %    97.5 %
(Intercept) 0.1003340 0.9436648
dect1       0.1947759 2.3412783
dect2       0.3807324 3.9010890
```

Examining the **OR** and the **CI**'s we see that the association is strongest when comparing the women who have had a mammogram with the last year `me=1`, to those who never had one, and comparing `dect=2` to `dect=0`, i.e. $\text{OR}_1(2,1) = 8{,}22$.

```
> coef(multi2)
  (Intercept)      dect1     dect2
1   -2.565201  0.7062589 2.1062453
2   -1.178617 -0.3926042 0.1977986
> exp(coef(multi2))
  (Intercept)     dect1    dect2
1  0.07690374 2.0263961 8.217330
2  0.30770391 0.6752959 1.218717
```

```
> exp(confint(multi2))
, , 1

                  2.5 %     97.5 %
(Intercept) 0.01005804  0.5880059
dect1       0.24245780 16.9360656
dect2       1.05675131 63.8982038

, , 2

                 2.5 %    97.5 %
(Intercept) 0.1003340 0.9436648
dect1       0.1947759 2.3412783
dect2       0.3807324 3.9010890
```

Examining the **OR** and the **CI**'s we see that the association is strongest when comparing the women who have had a mammogram with the last year `me=1`, to those who never had one, and comparing `dect=2` to `dect=0`, i.e. $\text{OR}_1(2,1) = 8{,}22$.

Interpretation: The odds of having a mammogram within the last year among the women who feel that a mammogram is very likely to detect a new case of breast cancer is $8{,}22$ times larger than the odds among women who feel that it is no likely.

Multinomial regression

- If the predictor is continuous, it will be modelled as linear in the logit, and will have a single parameter for each **logit** function
- This coefficient, when exponentiated, gives the estimated odds ratio (**OR**) for a change of one unit in the continuous variable.

Multinomial regression

Variables selection

- The steps to determine which covariates should be included in the model are similar to those in the logistic regression case.
- Let us consider the rest of variables in the study:

univariate/multivariate fits

```
> multi3 <- multinom(me~symp,data=mamexp); anova(multi0,multi3)
> multi4 <- multinom(me~bse,data=mamexp); anova(multi0,multi4)
> multi5 <- multinom(me~pb,data=mamexp); anova(multi0,multi5)
> multi6 <- multinom(me~symp+pb+bse+hist+dect,data=mamexp)
```

Multinomial regression

Variables selection

- The steps to determine which covariates should be included in the model are similar to those in the logistic regression case.
- Let us consider the rest of variables in the study:

univariate/multivariate fits

```
> multi3 <- multinom(me~symp,data=mamexp); anova(multi0,multi3)
> multi4 <- multinom(me~bse,data=mamexp); anova(multi0,multi4)
> multi5 <- multinom(me~pb,data=mamexp); anova(multi0,multi5)
> multi6 <- multinom(me~symp+pb+bse+hist+dect,data=mamexp)
```

vglm

```
>  vglm6 <- vglm(me ~ symp+pb+bse+hist+dect,
+                  family=multinomial(refLevel=1), data=mamexp)
```

- `vglm` function computes Wald statistics and p-values, `summary(vglm6)` gives:

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.99875    1.53905  -1.948  0.05136 .
(Intercept):2 -0.98609    1.11184  -0.887  0.37513
symp2:1        0.11004    0.92273   0.119  0.90508
symp2:2       -0.29008    0.64406  -0.450  0.65243
symp3:1        1.92471    0.77757   2.475  0.01331 *
symp3:2        0.81731    0.53979   1.514  0.12999
symp4:1        2.45699    0.77530   3.169  0.00153 **
symp4:2        1.13224    0.54767   2.067  0.03870 *
pb:1          -0.21944    0.07551  -2.906  0.00366 **
pb:2          -0.14821    0.07637  -1.941  0.05230 .
hist1:1        1.36624    0.43752   3.123  0.00179 **
hist1:2        1.06544    0.45940   2.319  0.02038 *
bse1:1         1.29167    0.52988   2.438  0.01478 *
bse1:2         1.05214    0.51499   2.043  0.04105 *
dect1:1        0.01702    1.16169   0.015  0.98831
dect1:2       -0.92439    0.71375  -1.295  0.19528
dect2:1        0.90414    1.12661   0.803  0.42225
dect2:2       -0.69053    0.68712  -1.005  0.31491
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

  - It suggests that with the possible exception of `dect`, each of the variables contribute to the model

- Let us look at the coefficients for the variable `symp`

```
Coefficients:
.......................................................
symp2:1        0.11004    0.92273   0.119  0.90508
symp2:2       -0.29008    0.64406  -0.450  0.65243
symp3:1        1.92471    0.77757   2.475  0.01331 *
symp3:2        0.81731    0.53979   1.514  0.12999
symp4:1        2.45699    0.77530   3.169  0.00153 **
symp4:2        1.13224    0.54767   2.067  0.03870 *
.......................................................
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

  - `symp2:1` and `symp2:2` estimates the log odds for `Agree` VS `Strongly Agree` (the reference value). This suggests that both categories are similar since the Wald statistics is significant.

- **Then is `symp` significant?**

- We can use a simpler model by coding `symp` into two levels: `0=Agree/Strongly agree` and `1=Disagree or Strongly disagree`

```
> mamexp$symp01<-mamexp$symp  # create a new variable symp01
> require(car)
> mamexp$symp01<-recode(mamexp$symp,"1=0;2=0;3=1;4=1")
```

  - Fit the model with `symp01` a look at the significance of the coefficients

```
>  vglm6a <- vglm(me ~ symp01+pb+bse+hist+dect,
+                  family=multinomial(refLevel=1), data=mamexp)
```

Multinomial regression

Variables selection

- Let us compare `multi6` with a model that excludes `dect`, and apply a LRT:

```
> multi6a <- multinom(me~symp01+pb+bse+hist,data=mamexp)
> multi7a <- multinom(me~symp01+pb+bse+hist+dect,data=mamexp)
> anova(multi6a,multi7a)


Likelihood ratio tests of Multinomial Models

Response: me
                             Model Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
1        symp01 + pb + bse + hist       814   706.0381
2 symp01 + pb + bse + hist + dect       810   697.4959 1 vs 2     4 8.542124 0.07362063
```

The LRT indicates that `dect` is not significant and should be removed from the model.

Multinomial regression

Variables selection

- Let us compare multi6 with a model that excludes dect, and apply a LRT:

```
> multi6a <- multinom(me~symp01+pb+bse+hist,data=mamexp)
> multi7a <- multinom(me~symp01+pb+bse+hist+dect,data=mamexp)
> anova(multi6a,multi7a)


Likelihood ratio tests of Multinomial Models

Response: me
                            Model Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
1         symp01 + pb + bse + hist       814   706.0381
2 symp01 + pb + bse + hist + dect       810   697.4959 1 vs 2     4 8.542124 0.07362063
```

The LRT indicates that dect is not significant and should be removed from the model.
But can we try something such as re-group the categories of dect?

- Let us explore collapsing `dect` into two categories: `0=NotLikely/SomewhatLikely` and `1=VeryLikely`

- Let us explore collapsing `dect` into two categories: `0=NotLikely/SomewhatLikely` and `1=VeryLikely`

```
> table(dect)

dect
  0   1   2
 18 105 289

> mamexp$dect01 <- mamexp$dect
> mamexp$dect01 <- recode(mamexp$dect01,"0=0; 1=0; 2=1")
```

- Let us explore collapsing `dect` into two categories: `0=NotLikely/SomewhatLikely` and `1=VeryLikely`

```
> table(dect)

dect
  0   1   2
 18 105 289

> mamexp$dect01 <- mamexp$dect
> mamexp$dect01 <- recode(mamexp$dect01,"0=0; 1=0; 2=1")
```

  - Fit the model with `symp01` and `dect01` a look at the significance of the coefficients

```
>  vglm7a <- vglm(me ~ symp01+pb+bse+hist+dect01,
+                family=multinomial(refLevel=1), data=mamexp)
```

Can we use `dect01` in our model?

Multinomial regression

Variables selection

- Now, we consider the variable `pb` (continuous and discrete) which is the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.

[1]We cannot use the function `gam` as we did before, as the multinomial family is not available in `mgcv`

Multinomial regression

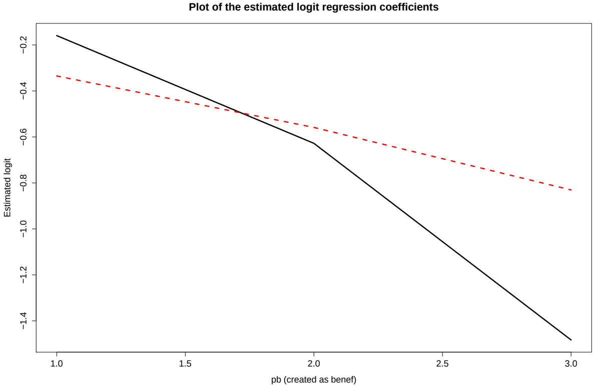
Variables selection

- Now, we consider the variable pb (continuous and discrete) which is the sum of five scaled responses, each on a four point scale. A low value is indicative of a woman with strong agreement with the benefits of mammography.
- Now, we want to know if the relationship of pb and the logit is linear or not[1]
- Let us recode the variable pb, range of pb is [5,17], into

```
> mamexp$benef <- mamexp$pb
> mamexp$benef[mamexp$benef<=5] =  0
> mamexp$benef[mamexp$benef>5 & mamexp$benef <=7] =  1
> mamexp$benef[mamexp$benef>7 & mamexp$benef <=9] =  2
> mamexp$benef[mamexp$benef>9] = 3
> mamexp$benef <- factor(mamexp$benef)
```

[1] We cannot use the function gam as we did before, as the multinomial family is not available in mgcv

- benef is a factor variable created to evaluate the scale of the effect of pb
- We plot the estimated logits with the design variable benef
- See Multinom.R script



- The plot shows evidence of linearity in the logits in pb, and hence we can leave pb as linear continuous predictor.

Multinomial regression

Variables selection

- As in the case of logistic regression, the next step is to assess the need to include interaction terms in the model.
- In this case, none of the possible interactions are significant.
- Our preferred model would be `vglm7a` or equivalently with `multinom()`

```
> multi9 <- multinom(me~hist+symp01+bse+pb+dect01,data=mamexp)
> exp(coef(multi9))
  (Intercept)    hist1  symp011     bse1        pb  dect011
1  0.07252958 3.705677 8.123425 3.445302 0.7792099 2.423432
2  0.16139719 2.895563 3.087680 2.601307 0.8569992 1.120929
```

Multinomial regression

Variables selection

- As in the case of logistic regression, the next step is to assess the need to include interaction terms in the model.
- In this case, none of the possible interactions are significant.
- Our preferred model would be `vglm7a` or equivalently with `multinom()`

```
> multi9 <- multinom(me~hist+symp01+bse+pb+dect01,data=mamexp)

> exp(coef(multi9))

  (Intercept)    hist1  symp011     bse1        pb  dect011
1  0.07252958 3.705677 8.123425 3.445302 0.7792099 2.423432
2  0.16139719 2.895563 3.087680 2.601307 0.8569992 1.120929
```

- **How do you interpret the OR?**

Multinomial regression

interpretation

```
> exp(coef(multi9))

  (Intercept)    hist1  symp011      bse1        pb  dect011
1  0.07252958 3.705677 8.123425 3.445302 0.7792099 2.423432
2  0.16139719 2.895563 3.087680 2.601307 0.8569992 1.120929
```

Multinomial regression

interpretation

```
> exp(coef(multi9))

  (Intercept)    hist1  symp011     bse1        pb  dect011
1  0.07252958 3.705677 8.123425 3.445302 0.7792099 2.423432
2  0.16139719 2.895563 3.087680 2.601307 0.8569992 1.120929
```

- The estimated OR's show that generally, it increases from `"1=Within a year"` to `"2=Over a year"` wrt to the baseline category `"0=Never"`, except on the variable `pb`.
- The coefficients of the variable `pb` are negative (and hence OR$<$1). It suggests that larger values indicate less belief in the benefit of mammography screening.
- Hence, the OR for `pb` reflect that **"less belief"** is significantly associated with less frequent use.

Multinomial regression

interpretation

- The function `effect` in `library(effects)` construct terms effects for a multinomial model

```
> library(effects)
> plot(effect("symp01",multi9))
> plot(effect("hist",multi9))
> plot(effect("bse",multi9))
> plot(effect("dect01",multi9))
> plot(effect("pb",multi9),style="stacked")
> plot(effect("pb",multi9),style="lines")
```

Multinomial regression

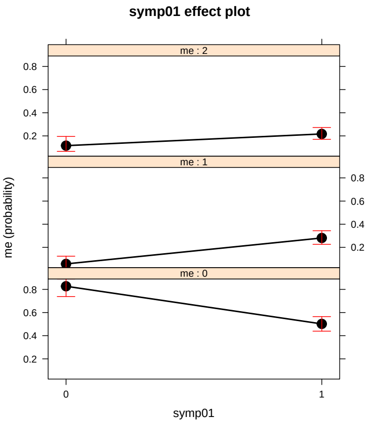
interpretation

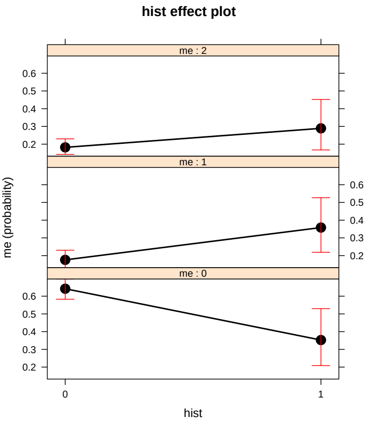
- The function `effect` in `library(effects)` construct terms effects for a multinomial model

```
> library(effects)
> plot(effect("symp01",multi9))
> plot(effect("hist",multi9))
> plot(effect("bse",multi9))
> plot(effect("dect01",multi9))
> plot(effect("pb",multi9),style="stacked")
> plot(effect("pb",multi9),style="lines")
```

Multinomial regression

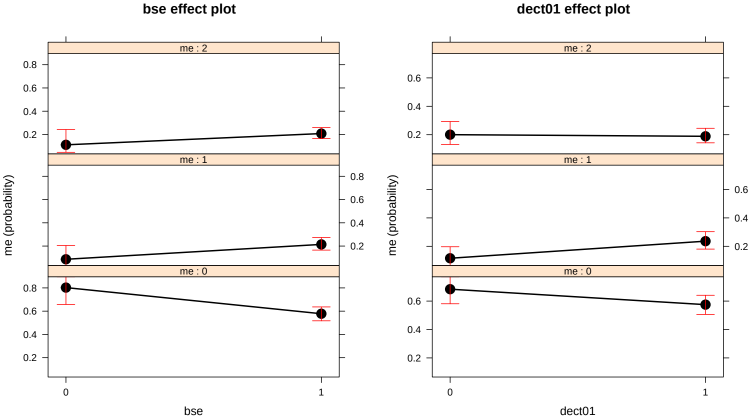
Effects plots





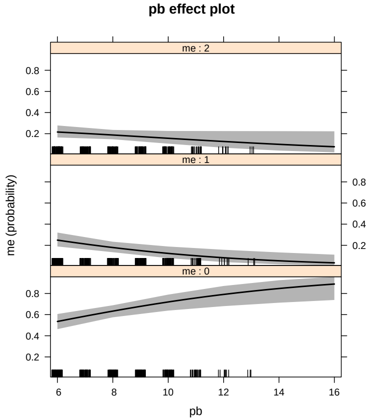
Multinomial regression

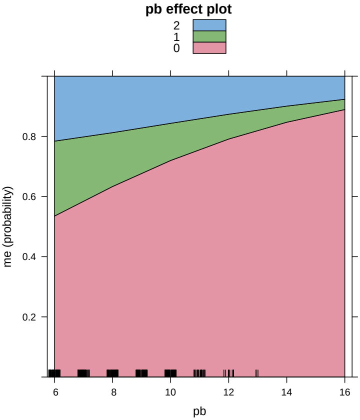
Effects plots



Multinomial regression

Effects plots





Multinomial regression

Conclusions

- The real challenge when fitting a multinomial logistic regression model is the fact that there are multiple odds ratios for each model covariate.
- This complicates the interpretation of the model.
- However, using a **multinomial outcome** provide more complete description of the process being studied.
- For instance, in the mammography experience study, if we had combined the outcome into a binary response (e.g.: `ever` VS `never`) we would have missed the gradation in odds ratios.
- In practice, one should not pool the outcome categories unless the estimated coefficients in the logits are not significantly different from each other.
- In summary, fitting and interpreting the results from a multinomial logistic regression model follows the same paradigm as in the binary logistic model.