

Background

Multinomial models are linear statistical models for which the response variable is a factor with over two levels. These models (also termed as generalized logit models) are extensions to the more familiar binomial logistic regression or logit models. The independent variables can be dichotomous or continuous. Similar to binary logistic regression, multinomial logistic regression uses maximum likelihood estimation to test the probability of categorical membership.

Assume that Y is a categorical variable with levels A, B, and C, and that there are m predictors $X^{(1)}, X^{(2)}, \dots, X^{(m)}$. Define the following quantities:

$$\begin{aligned}p_A &= P[Y = A|X] \\p_B &= P[Y = B|X] \\p_C &= P[Y = C|X]\end{aligned}$$

A multinomial logistic regression model has the following form:

$$\begin{aligned}\ln\left[\frac{p_A}{p_C}\right] &= \beta_{0,1} + \beta_{1,1} \cdot X^{(1)} + \beta_{2,1} \cdot X^{(2)} + \dots + \beta_{m,1} \cdot X^{(m)} \\ \ln\left[\frac{p_B}{p_C}\right] &= \beta_{0,2} + \beta_{1,2} \cdot X^{(1)} + \beta_{2,2} \cdot X^{(2)} + \dots + \beta_{m,2} \cdot X^{(m)}\end{aligned}$$

- Since $p_A + p_B + p_C = 1$, if $\ln\left[\frac{p_A}{p_C}\right]$ and $\ln\left[\frac{p_B}{p_C}\right]$ are known, then all three probabilities can be calculated.
- If the response variable has k classes, then the multinomial logistic regression model will need to consist of k-1 equations.
- The expression p_Y/p_X is referred to as the relative odds ratio of Y with respect to X.

Example

To demonstrate this analysis data set utilized is a frequency table from a Copenhagen housing conditions survey from the `MASS` library. The `housing` data is transformed into 1681 rows and 5 variables.

- Sat - Satisfaction of householders with their present housing circumstances, (High, Medium or Low, ordered factor).
- Infl - Perceived degree of influence householders have on the management of the property (High, Medium, Low).
- Type - Type of rental accommodation, (Tower, Atrium, Apartment, Terrace).
- Cont - Contact residents are afforded with other residents, (Low, High).
- Freq - Frequencies: the numbers of residents in each class.

```
library(MASS)
library(tidyverse)
library(nnet)
library(effects)

t = housing$Freq
housing.df = as.data.frame(apply(housing, 2, function(x) rep(x, t)), row.names = FALSE)
head(housing.df) %>% knitr::kable(caption = "Conditions of Copenhagen Housing")
```

Conditions of Copenhagen Housing

Sat	Infl	Type	Cont	Freq
Low	Low	Tower	Low	21
Low	Low	Tower	Low	21
Low	Low	Tower	Low	21
Low	Low	Tower	Low	21

Sat	Infl	Type	Cont	Freq
Low	Low	Tower	Low	21
Low	Low	Tower	Low	21

```
# Format categorical variables
housing.df$Sat = factor(housing.df$Sat, levels = c("Low", "Medium", "High"))
housing.df$Infl = factor(housing.df$Infl, levels = c("Low", "Medium", "High"))
housing.df$Type = factor(housing.df$Type)
housing.df$Cont = factor(housing.df$Cont, levels = c("Low", "High"))
str(housing.df)
```

```
## 'data.frame': 1681 obs. of 5 variables:
## $ Sat : Factor w/ 3 levels "Low","Medium",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Infl: Factor w/ 3 levels "Low","Medium",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Type: Factor w/ 4 levels "Apartment","Atrium",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Cont: Factor w/ 2 levels "Low","High": 1 1 1 1 1 1 1 1 1 1 ...
## $ Freq: chr "21" "21" "21" "21" ...
```

```
summary(housing.df)
```

```
##      Sat      Infl      Type      Cont      Freq
## Low   :567 Low   :627 Apartment:765 Low :713 Length:1681
## Medium:446 Medium:659 Atrium   :239 High:968 Class :character
## High  :668 High   :395 Terrace :277           Mode  :character
##                               Tower  :400
```

Building the Model

Multinomial logistic regression reports the odds of being in the different outcome categories about some base groups. A model is built to capture the odds of the satisfaction of householders with their present housing circumstances, `Sat`, based on the independent variables. Because there are three levels to `Sat`, the model will report two distinct sets of regression results corresponding to the following two models.

$$\log\left(\frac{Pr(Sat = Medium)}{Pr(Sat = Low)}\right) = \beta_0 + \beta_1(Infl) + \beta_2(Type) + \beta_3(Cont) + \varepsilon$$

$$\log\left(\frac{Pr(Sat = High)}{Pr(Sat = Low)}\right) = \beta_0 + \beta_1(Infl) + \beta_2(Type) + \beta_3(Cont) + \varepsilon$$

In this case, the “Low” satisfaction is treated as the reference group. Now, using the `multinom()` command from the `nnet` package, a multinomial log-linear models via neural networks is fitted.

```
model = multinom(Sat ~ Infl + Type + Cont, data = housing.df)
```

```
## # weights: 24 (14 variable)
## initial value 1846.767257
## iter 10 value 1744.380656
## final value 1735.041933
## converged
```

```
summary(model)
```

```
## Call:
## multinom(formula = Sat ~ Infl + Type + Cont, data = housing.df)
##
## Coefficients:
##      (Intercept) InflMedium InflHigh TypeAtrium TypeTerrace TypeTower
## Medium -0.8549167  0.4463953 0.6649354  0.5670590 -0.2308827 0.4356897
## High   -0.8743756  0.7348634 1.6126341  0.3276536 -0.6766988 0.7356326
##      ContHigh
## Medium 0.3608516
## High   0.4818282
##
## Std. Errors:
##      (Intercept) InflMedium InflHigh TypeAtrium TypeTerrace TypeTower
## Medium 0.1490307  0.1415573 0.1863376  0.1946515  0.1748384 0.1725329
## High    0.1437481  0.1369380 0.1671318  0.1886343   0.1755643 0.1552715
##      ContHigh
## Medium 0.1323976
## High    0.1241371
##
## Residual Deviance: 3470.084
## AIC: 3498.084
```

The model-running output includes some iteration history, and the final negative log-likelihood is 1735.04. This value multiplied by two is the residual deviance and it can be used in comparisons of nested models.

The summary results in the two models when “Low” is the reference point. In other words, the rows with `Medium` are for the model comparing the probability of having a “Medium” satisfaction level versus a “Low” satisfaction level. While the rows with `High` are for the model comparing the probability of having a “High” satisfaction level versus a “Low” satisfaction level.

Next, p-value calculation for the regression coefficients is done using Wald tests to determine whether coefficients are significant or not at $\alpha = 0.05$.

```
z = summary(model)$coefficients/summary(model)$standard.errors
(1 - pnorm(abs(z), 0, 1)) * 2
```

```
##      (Intercept) InflMedium InflHigh TypeAtrium TypeTerrace
## Medium 9.664561e-09 1.613470e-03 0.0003591066 0.003577437 0.1866516550
## High   1.181792e-09 8.032596e-08 0.0000000000 0.082391184 0.0001160037
##      TypeTower ContHigh
## Medium 1.156140e-02 0.0064201178
## High   2.161372e-06 0.0001038484
```

Model Plots & Interpretation

Thus, the final model becomes:

$$\log\left(\frac{\Pr(\hat{Sat} = \text{Medium})}{\Pr(Sat = \text{Low})}\right) = -0.85 + 0.46(\text{Infl}_{\text{Medium}}) + 0.66(\text{Infl}_{\text{high}}) + 0.57(\text{Type}_{\text{Atrium}}) + 0.44(\text{Type}_{\text{Tower}}) + 0.36(\text{Cont}_{\text{High}})$$

$$\log\left(\frac{\Pr(\hat{Sat} = \text{High})}{\Pr(Sat = \text{Low})}\right) = -0.87 + 0.73(\text{Infl}_{\text{Medium}}) + 1.61(\text{Infl}_{\text{high}}) - 0.68(\text{Type}_{\text{Terrace}}) + 0.74(\text{Type}_{\text{Tower}}) + 0.48(\text{Cont}_{\text{High}})$$

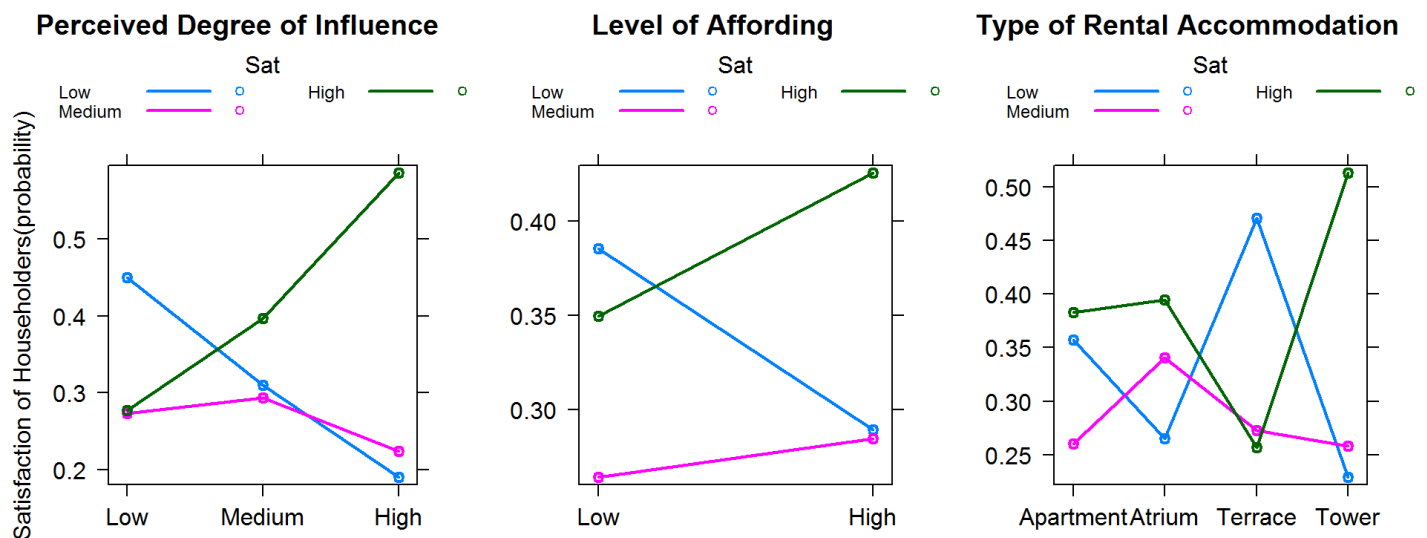
where

- The *log-odds* for ‘Medium’ satisfaction vs. ‘Low’ satisfaction will have:
 - Infl*: increase by 0.46 if moving from “low” to “medium”, and by 0.66 if moving from “low” to “high” in the perceived degree of influence.
 - Type*: increase by 0.57 if moving from “Apartment” to “Atrium”, and by 0.44 if moving from “Apartment” to “Tower” based on the type of rental accommodation.
 - Cont*: increase by 0.36 if moving from “low” to “high” in that contact residents are afforded with other residents.
- The *log-odds* for ‘High’ satisfaction vs. ‘Low’ satisfaction will have:
 - Infl*: increase by 0.73 if moving from “low” to “medium”, and by 1.61 if moving from “low” to “high” in the perceived degree of influence.

- *Type*: decreases by 0.68 if moving from “Apartment” to “Terrace”, and increases by 0.74 if moving from “Apartment” to “Tower” based on the type of rental accommodation.
- *Cont*: increase by 0.48 if moving from “low” to “high” in that contact residents are afforded with other residents.

The plots below highlight the effect of each predictor according to their change in factors. For instance, there is a small difference in medium and high satisfaction levels when the perceived degree of influence is low. Satisfaction is higher when the perceived degree of influence is high.

```
p1 = plot(Effect("Infl", model), multiline = TRUE, axes=list(
  x = list(Infl = list(lab = "")),
  y = list(lab = "Satisfaction of Householders(probability)"),
  main = "Perceived Degree of Influence")
p2 = plot(Effect("Type", model), multiline = TRUE, axes=list(
  x = list(Type = list(lab = "")),
  y = list(lab = "")),
  main = "Type of Rental Accommodation")
p3 = plot(Effect("Cont", model), multiline = TRUE, axes=list(
  x = list(Cont = list(lab = "")),
  y = list(lab = "")),
  main = "Level of Affording")
gridExtra::grid.arrange(p1,p3,p2,nrow =1)
```



Predictions

With the use of the model, predictions can be made defining the levels for the predictors. For example, let's assume the satisfaction of an apartment householder with their present housing circumstances is needed based on low perceived degree of influence householder have on the management of the property, and contact residents are afforded with other residents is high.

```
predict(model, data.frame(Type = "Apartment", Infl = "Low", Cont = "High"))
```

```
## [1] Low
## Levels: Low Medium High
```

As a result, the satisfaction of the householder with their present housing circumstances is expected to be 'low'.

Works Cited

- Multinomial Logistic Regression. UCLA: Statistical Consulting Group. from <https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/> (https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/) (accessed October 12, 2020)