# ***Assignment 2***

The *binary choice* and its application on an example will be examined in this analysis. In statistics, the (binary) logistic model (or logit model) is a statistical model that models the probability of one event (out of two alternatives) taking place by having the log-odds (the logarithm of the odds) for the event be a linear combination of one or more independent variables ("predictors").The CarBin data will be used,which are in wide format, which contains the following information:

1. Choice: choice of a vehicle among 2 propositions
2. College: college education
3. Hsg2: size of household greater than 2
4. Coml5: commute lower than 5 miles a day
5. Pricez: price divided by log
6. Typez : body type, one of regcar (regular car), sportuv (sport utility vehicle), sportcar, stwagon (station wagon), truck, van, for each alternative z (1 or 2)
7. rangez: hundreds of miles vehicle can travel between refuelings/rechargings
8. accelz: acceleration, tens of seconds required to reach 30 mph from stop
9. speedz: highest attainable speed in hundreds of mph
10. pollutez: tailpipe emissions as fraction of those for new gas vehicle
11. sizez: 0 for a mini, 1 for a subcompact, 2 for a compact and 3 for a mid–size or large vehicle
12. costz: cost per mile of travel (tens of cents) : home recharging for electric
13. vehicle, station refueling otherwise.

What we can see is that the data frame is unlabeled because our options are supplied as numbers 1 and 2 without any explanation as to what these two numbers mean. Following that, we'll estimate a model containing demographics and all alternative-specific characteristics having generic parameters. We used the price, range, speed, cost, station, pollution, accel, and some variable dummies we constructed as generic parameters in the model (typetvan, typestwagon, typetruck, typesportuv, size0, size1, size2) and as alternative specific parameters, the hsg2, coml5, and college variables.

| ***Coefficients*** | ***Estimate*** | ***Pr(>\|z\|)*** |
|---|---|---|
| **Intercept** | -1.35122429 | 5.11E-10 |
| **Price** | -0.19040515 | 0.0076982 |
| **Range** | 0.00439735 | 4.44E-16 |
| **Speed** | 0.00452134 | 0.0385467 |
| **Cost** | -0.05386259 | 0.0054712 |
| **Accel** | -0.05999582 | 0.0350559 |
| **Typewagon** | -0.39971949 | 0.0326645 |
| **Typetruck** | -0.47815884 | 0.0005424 |
| **Hsg2:2** | 0.61173305 | 5.33E-05 |

Table 1.Formula 1 -Coefficient estimation

Table 1 shows the statistically significant coefficients and shows which have a negative association and which have a positive link depending on their sign. Choice 1 has been used as a reference level in this case. Then a model with demographics and some alternative-specific characteristics with alternative-specific parameters will be estimated.

| ***Coefficients*** | ***Estimate*** | ***Pr(>\|z\|)*** |
|---|---|---|
| **Price** | -0.18648958 | 0.009235 |
| **Range** | 0.00435723 | 1.11E-15 |
| **Cost** | -0.05363336 | 0.0054487 |
| **Typestwagon** | -0.40925557 | 0.02795 |
| **Typetruck** | -0.48328019 | 0.0004925 |
| **Hsg2:2** | 0.60323945 | 6.71E-05 |
| **Accel:2** | -0.13088251 | 0.0062611 |

Table 2.Formula 2-Coefficient estimation

Table 2 provides the statistically significant rates and illustrates which have a negative and which have a positive association based on their sign, with Choice 1 serving as a reference level. However, there are some distinctions between this model and the previous one. Some generic characteristics in formula 2 have been relocated to alternative specifics. These characteristics are speed, pollution, and acceleration.

Then we'll compare Model 2 to Model 1. This will be accomplished through the use of the Likelihood Ratio Test.

| *Model* | *LogLik* | *Chisq* | *Pr(Chisq)* |
|---|---|---|---|
| **1** | -667.79 | | |
| **2** | -666.97 | 1.6536 | < 2.2e-16 |

Table 3.Likelihood Ratio Test

The logL of model 1 is -667.79 and the logL of model 2 is -666.97.The test statistic is twice the difference in lnL: 2*(667.79-666.97) =1.6536.So therefore we compare 1.6536 with the critical value of chi-squared .The null hypothesis is that the model 1 is the “true” model, with a large test statistic we reject null hypothesis .In this case we reject null hypothesis because P-value< 2.2e-16. Is not the expected outcome because the data is unlabeled so it is assumed that no alternative specific parameters need to be created.

| *Attribute* | *WTP* |
|---|---|
| **Intercept** | -7.309545 |
| **Range** | 0.02288451 |
| **Speed** | 0.02377382 |
| **Cost** | -0.2815401 |
| **Station** | 1.64237 |
| **Pollute** | 0.7731793 |
| **Accel** | -0.3160867 |
| **Typevan** | 0.2062025 |
| **Typestwagon** | -1.970342 |
| **Ttypetruck** | -2.629901 |
| **Typesportuv** | -1.529233 |
| **Typesportcar** | -2.494176 |
| **Size0** | -1.732076 |
| **Size1** | -1.275185 |
| **Size2** | 0.007772812 |

Table 4 .WTP for each attribute

The Willingness To-Pay for each attribute is shown in table 4. *Willingness to pay*, sometimes abbreviated as WTP, is the maximum price a customer is willing to pay for a product or service. When the value is positive, the respondent is willing to pay the amount indicated for each attribute, however when the price is negative is the opposite.

| *Minimum* | *Median* | *Mean* | *Maximum* |
|---|---|---|---|
| -3.2466 | -0.8214 | -0.856 | -0.1522 |

Table 5.Welfare Effect

What is shown in table 5 is those with the shift in pollute and price, it is as if -0.8560 (mean) was taken from the respondents.

# Code

```
library("mlogit")
library("dfidx")
library("Formula")
library("lmtest")
car= read.csv("CarBin.csv", header=TRUE,sep = ';')
carbin<-dfidx(car, varying = 5:22,sep = '',
        choice = "choice",
        idnames=c("chid","alt")
        )
table(carbin$size)
carbin$size0=ifelse(carbin$size==0,1,0)
carbin$size1=ifelse(carbin$size==1,1,0)
carbin$size2=ifelse(carbin$size==2,1,0)
carbin$size3=ifelse(carbin$size==3,1,0)
carbin$typeregcar=ifelse(carbin$type=="regcar",1,0)
carbin$typesportuv=ifelse(carbin$type=="sportuv",1,0)
carbin$typestwagon=ifelse(carbin$type=="stwagon",1,0)
carbin$typetruck=ifelse(carbin$type=="truck",1,0)
carbin$typetvan=ifelse(carbin$type=="van",1,0)
carbin$typesportcar=ifelse(carbin$type=="sportcar",1,0)

myformula1=Formula(choice ~ price+range+speed+cost+station+pollute+accel
            +typetvan+typestwagon+typetruck+typesportuv+size0+size1+typesportcar
            +size2|hsg2+coml5+college|0)
myformula2=Formula(choice~price+range+cost+station+typesportcar
```

```
          +typetvan+typestwagon+typetruck+typesportuv+size0+size1
          +size2|hsg2+coml5+college+speed+pollute+accel)
ml.tm1=mlogit(myformula1,carbin,reflevel ="1")
ml.tm2=mlogit(myformula2,carbin,reflevel ="1")
ml.tm1
summary(ml.tm1)
summary(ml.tm2)
lrtest(ml.tm2,ml.tm1)
-coef(ml.tm1)[1]/coef(ml.tm1)[2]
-coef(ml.tm1)[3]/coef(ml.tm1)[2]
-coef(ml.tm1)[4]/coef(ml.tm1)[2]
-coef(ml.tm1)[5]/coef(ml.tm1)[2]
-coef(ml.tm1)[6]/coef(ml.tm1)[2]
-coef(ml.tm1)[7]/coef(ml.tm1)[2]
-coef(ml.tm1)[8]/coef(ml.tm1)[2]
-coef(ml.tm1)[9]/coef(ml.tm1)[2]
-coef(ml.tm1)[10]/coef(ml.tm1)[2]
-coef(ml.tm1)[11]/coef(ml.tm1)[2]
-coef(ml.tm1)[12]/coef(ml.tm1)[2]
-coef(ml.tm1)[13]/coef(ml.tm1)[2]
-coef(ml.tm1)[14]/coef(ml.tm1)[2]
-coef(ml.tm1)[15]/coef(ml.tm1)[2]
-coef(ml.tm1)[16]/coef(ml.tm1)[2]
carbin2=carbin
carbin2$pollute=carbin$pollute*0.8
carbin2$price=carbin2$price*1.2
```

```
ivbefore=logsum(ml.tm1)

ivafter=logsum(ml.tm1,data=carbin2)

surplus=-(ivafter-ivbefore)/coef(ml.tm1)["price"]

summary(surplus)
```