lab7

Logistical regression

At first we read the data from `survey.csv`

```
dat <- read.csv("survey.csv",header=TRUE,sep=",")
```

Then we slightly look around our data

```
head(as.factor(dat$Price))
```

```
## [1] 10 20 30 20 30 20
## Levels: 10 20 30
```

```
table(dat$MYDEPV)
```

```
##
##   0   1
## 426 324
```

```
with(dat, table(Price,MYDEPV))
```

```
##      MYDEPV
## Price   0   1
##    10 115 135
##    20 137 113
##    30 174  76
```

Make first log regression that predict `MYDEPV` with `Income`, `Age` and `Price`:

```
logModel <- glm(MYDEPV ~ Income + Age + as.factor(Price),
               data = dat,
               family=binomial(link="logit"),
               na.action=na.pass)
summary(logModel)
```

```
##
## Call:
## glm(formula = MYDEPV ~ Income + Age + as.factor(Price), family = binomial(link = "logit"),
##     data = dat, na.action = na.pass)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0388  -0.5581  -0.2434   0.4178   3.2377
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -6.02116    0.53244 -11.309  < 2e-16 ***
## Income               0.12876    0.00923  13.950  < 2e-16 ***
## Age                  0.03506    0.01179   2.974  0.00294 **
## as.factor(Price)20 -0.74418    0.26439  -2.815  0.00488 **
## as.factor(Price)30 -2.21028    0.31108  -7.105  1.2e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1025.81  on 749  degrees of freedom
## Residual deviance:  534.17  on 745  degrees of freedom
## AIC: 544.17
##
## Number of Fisher Scoring iterations: 6
```

```
confint(logModel)
```

```
## Waiting for profiling to be done...
```

```
##                          2.5 %      97.5 %
## (Intercept)        -7.10166938 -5.01101651
## Income              0.11149784  0.14774622
## Age                 0.01197129  0.05830305
## as.factor(Price)20 -1.26927449 -0.23079467
## as.factor(Price)30 -2.84054121 -1.61841547
```

Coefficients Analysis

For every change by one unit result will change on such percent:

```
100 * exp(logModel$coefficients) - 100
```

```
##        (Intercept)             Income                Age
##         -99.757315          13.741640           3.568576
## as.factor(Price)20 as.factor(Price)30
##         -52.487507         -89.033011
```

Check sums

Test the rule that the probability mass equals the counts:

```
dat$predicted <- predict(logModel,newdata=data.frame(Income=dat$Income,Age=dat$Age,Price=dat$Price),typ
sum(dat$predicted)
```

```
## [1] 324
```

```
sum(dat$MYDEPV)
```

```
## [1] 324
```

They are equal! Great

Test!

```
testData <- data.frame (Age = c(99), Income = c(36.572), Price = c(30))
testData$Prob <- predict(logModel,newdata=testData,type="response")
testData
```

```
##   Age Income Price     Prob
## 1  99 36.572    30 0.487218
```