Tutorial Business Analytics

Homework 4

Homework 4.1

<u>Note:</u> Use R to solve this exercise (Homework4.1_R-template.R).

Install/open the “AER” (Applied Econometrics with R) package and open the “RecreationDemand” data set.

a) Briefly describe the data set:

   i. Name the dependent variable and the independent variables.
   ii. Which scales of measurement do the variables belong to (e.g. nominal, ordinal, interval or ratio)?
   iii. Does the data set consist of cross-sectional, time-series or panel data?

Estimate a Poisson Regression (rd_pois), in which the number of boat trips is regressed against all explanatory variables:

$$\begin{aligned} ln\big(\mu(trips_i)\big) = \beta_0 + \beta_1 \cdot income_i + \beta_2 \cdot costC_i + \beta_3 \cdot costS_i + \beta_4 \cdot costH_i + \beta_5 \cdot ski_i \\ + \beta_6 \cdot userfee_i + \beta_7 \cdot quality_i \end{aligned}$$

```
rd_pois <- glm(trips~income+userfee+costC+costS+costH+ski+quality, data =
RecreationDemand, family = poisson)

summary(rd_pois)
```

Be aware that quality is treated as a numeric variable in our model, although it is an ordinal variable (s. data set description for reasons).

b) The Poisson distribution has the property that variance equals mean (equidisperision). Thus, the Poisson Regression can only be applied if the mean of boat trips equals its variance. Is the equidispersion assumption fulfilled in our data? Use the “Dispersion Test” to give an answer:

```
dispersiontest(rd_pois)
```

c) As a consequence of overdispersion you decide to reestimate the above model, but using a Negative Binomial Regression:

   First install/open the “MASS” (Modern Applied Statistics with S) package:

```
library("MASS")    #install.packages("MASS")
```

Then estimate:

```
rd_nb<- glm.nb(trips~income+userfee+costC+costS+costH+ski+quality, data =
RecreationDemand)

summary(rd_nb)
```

d) Which attributes are statistically significant regarding a significance level of 5%?

e) Interpret the coefficients.

Solution - Homework 4.1

a)

i. dependent variable: trips (boat trips in 1980)
   independent variables: quality, ski, income, userfee, costC, costS, costH

ii. ratio: trips, income, costC/S/H
    nominal: ski, userfee
    ordinal: quality

Be careful, although quality is an ordinal variable, it is treated as a numeric variable (s. data set description)!!!

iii. cross-sectional: 2000 boat owners in 1980

b) to e) s. Homework 4.1_R-Script

Homework 4.2

You want to examine the relationship between age and owning a car. Owning a car is modeled as a binary variable, taking on the value of one when true and zero if not. Therefore you employed a logistic regression and obtained the following results:

Variable	Est. coefficient	Standard error
Age	0.135	0.036
Constant	-3.89	1.73

a) According to the model above, what effect (qualitative and quantitative) does a change in age (+1) have on the dependent variable?

b) Find the age for which the model would be indifferent between owning a car and not owning one.
An approximated solution using R’s plotting functionality is sufficient.

Solution - Homework 4.2

a)

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 = -3.89 + 0.135 \cdot age$$

$$e^{0.135} = 1.145 = \frac{\frac{p(age+1)}{1-p(age+1)}}{\frac{p(age)}{1-p(age)}} \equiv \frac{odds(age+1)}{odds(age)} = OR \quad \text{(Odds Ratio)}$$

$$\frac{p(age+1)}{1-p(age+1)} = 1.145 \cdot \frac{p(age)}{1-p(age)}$$

The odds of having a car (instead of not having a car) increase by 14.5% if the independent variable age gains one unit (unit = year in this case). In other words, it becomes 14.5% more likely to own a car relative to not owning one.

It follows that the probability of owning a car increases as well, but an exact number cannot be determined.

b)

<u>Indifference ($x_1{}^*$):</u> $\quad p = 1 - p = 0.5$

$$<=> \frac{p}{1-p} = \frac{0.5}{0.5} = 1$$

$$=> \ln\left(\frac{p}{1-p}\right) = \ln(1) = 0$$

$$=> \beta_0 + \beta_1 x_1{}^* = 0 \qquad <=> x_1{}^* = -\frac{\beta_0}{\beta_1} \text{ (See lecture slides)}$$

$$<=> -3.89 + 0.135 * age^* = 0 \qquad <=> age^* = \frac{3.89}{0.135} = 28.8$$

28.8 years is the age, at which the model would be indifferent between alternative 1 and 2.

You could also plot this in R to get an approximate solution. Compute the values of *p* for different values of *age* and then plot the result.

Recall that,

$$p = \frac{e^{-3.89+(0.135\times age)}}{1+e^{-3.89+(0.135\times age)}}$$

```
age = seq(from=0, to=60, by=1)
p = (exp(-3.89+(0.135*age)))/(1+(exp(-3.89+(0.135*age))))
plot(p, age)

# A more elaborate example using ggplot
# p_c = 0.5
# age_c = (log(p_c/(1-p_c))+3.89)/0.135
# ggplot(data=data.frame(age, p), aes(p, age)) + geom_point(size=1, color="blue") +
geom_point(data=data.frame(age=age_c,p=p_c), color="red", size=2)
```