Group_2_Analysis

Group 2

2022/3/20

```
library(tidyverse)
library(moderndive)
library(gapminder)
library(sjPlot)
library(stats)
library(jtools)
library(MASS)
library(kableExtra)
library(olsrr)
library(qcc)
```

```
#import data
data<-read.csv("dataset2.csv")

#processing discrete data
data[, 4] <- as.factor(data[, 4])
data[, 6] <- as.factor(data[, 6])
data[, 11] <- as.factor(data[, 11])

data = data[, -2]
```

Introduction

Exploratory Data Analysis

Modelling and Results

Because the dependent variable of the data of this fitting model is the counting variable (the total number of families), and the independent variable is the continuity or category variable. In addition, the variable data are measured every three years, and the length of the whole observation concentration is unchanged. This study decided to use Poisson regression to fit the model. Poisson regression mainly has two assumptions. Firstly, the human time risk of different objects with the same characteristics and at the same time is homogeneous. Secondly, when the sample size is larger and larger, the mean of frequency tends to variance.

Fitting model

Preliminary fitting model

```
model<-glm(Total.Number.of.Family.members~Total.Household.Income+Total.Food.Expenditure+Household.Head.S
summary(model)
```

```
Call:
glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + Household.Head.Sex + Household.Head.Age +
    Type.of.Household + House.Floor.Area + House.Age + Number.of.bedrooms +
    Electricity, family = "poisson", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.6392  -0.6578  -0.1209   0.5018   2.7098

Coefficients:
                                                        Estimate Std. Error
(Intercept)                                            1.671e+00  8.230e-02
Total.Household.Income                                -4.266e-07  7.596e-08
Total.Food.Expenditure                                 5.239e-06  4.066e-07
Household.Head.SexMale                                 2.418e-01  3.739e-02
Household.Head.Age                                    -5.818e-03  1.080e-03
Type.of.HouseholdSingle Family                        -3.732e-01  3.047e-02
Type.of.HouseholdTwo or More Nonrelated Persons/Members -5.036e-01  2.447e-01
House.Floor.Area                                      -9.056e-05  3.033e-04
House.Age                                             -2.451e-03  1.177e-03
Number.of.bedrooms                                    -2.366e-02  1.680e-02
Electricity1                                          -5.232e-02  4.048e-02
                                                      z value Pr(>|z|)
(Intercept)                                            20.299  < 2e-16 ***
Total.Household.Income                                 -5.616 1.96e-08 ***
Total.Food.Expenditure                                 12.886  < 2e-16 ***
Household.Head.SexMale                                  6.467 1.00e-10 ***
Household.Head.Age                                     -5.386 7.21e-08 ***
Type.of.HouseholdSingle Family                        -12.250  < 2e-16 ***
Type.of.HouseholdTwo or More Nonrelated Persons/Members  -2.058   0.0396 *
House.Floor.Area                                       -0.299   0.7653
House.Age                                              -2.082   0.0374 *
Number.of.bedrooms                                     -1.409   0.1589
Electricity1                                           -1.293   0.1961
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1373.63  on 1248  degrees of freedom
Residual deviance:  881.01  on 1238  degrees of freedom
AIC: 4931.9

Number of Fisher Scoring iterations: 4
```

The stepwise method was used to complete the screening of independent variables

```r
step(model)
```

```
Start:  AIC=4931.87
Total.Number.of.Family.members ~ Total.Household.Income + Total.Food.Expenditure +
    Household.Head.Sex + Household.Head.Age + Type.of.Household +
    House.Floor.Area + House.Age + Number.of.bedrooms + Electricity

                         Df Deviance    AIC
- House.Floor.Area        1   881.10 4930.0
- Electricity             1   882.67 4931.5
- Number.of.bedrooms      1   883.00 4931.9
<none>                        881.01 4931.9
- House.Age               1   885.41 4934.3
- Household.Head.Age      1   910.02 4958.9
- Total.Household.Income  1   916.63 4965.5
- Household.Head.Sex      1   924.80 4973.7
- Type.of.Household       2  1028.11 5075.0
- Total.Food.Expenditure  1  1033.71 5082.6

Step:  AIC=4929.96
Total.Number.of.Family.members ~ Total.Household.Income + Total.Food.Expenditure +
    Household.Head.Sex + Household.Head.Age + Type.of.Household +
    House.Age + Number.of.bedrooms + Electricity

                         Df Deviance    AIC
- Electricity             1   882.78 4929.6
<none>                        881.10 4930.0
- Number.of.bedrooms      1   883.59 4930.4
- House.Age               1   885.80 4932.7
- Household.Head.Age      1   910.07 4956.9
- Total.Household.Income  1   917.76 4964.6
- Household.Head.Sex      1   924.93 4971.8
- Type.of.Household       2  1028.11 5073.0
- Total.Food.Expenditure  1  1033.71 5080.6

Step:  AIC=4929.64
Total.Number.of.Family.members ~ Total.Household.Income + Total.Food.Expenditure +
    Household.Head.Sex + Household.Head.Age + Type.of.Household +
    House.Age + Number.of.bedrooms

                         Df Deviance    AIC
<none>                        882.78 4929.6
- Number.of.bedrooms      1   886.06 4930.9
- House.Age               1   888.38 4933.2
- Household.Head.Age      1   911.64 4956.5
- Total.Household.Income  1   919.96 4964.8
- Household.Head.Sex      1   927.56 4972.4
- Type.of.Household       2  1030.52 5073.4
- Total.Food.Expenditure  1  1033.99 5078.8

Call:  glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
```

```
    Total.Food.Expenditure + Household.Head.Sex + Household.Head.Age +
    Type.of.Household + House.Age + Number.of.bedrooms, family = "poisson",
    data = data)

Coefficients:
                                            (Intercept)
                                              1.636e+00
                                 Total.Household.Income
                                             -4.333e-07
                                 Total.Food.Expenditure
                                              5.211e-06
                                 Household.Head.SexMale
                                              2.441e-01
                                     Household.Head.Age
                                             -5.808e-03
                         Type.of.HouseholdSingle Family
                                             -3.739e-01
Type.of.HouseholdTwo or More Nonrelated Persons/Members
                                             -5.039e-01
                                              House.Age
                                             -2.707e-03
                                     Number.of.bedrooms
                                             -2.859e-02

Degrees of Freedom: 1248 Total (i.e. Null);  1240 Residual
Null Deviance:      1374
Residual Deviance: 882.8    AIC: 4930
```

Use a better model

```
model.best<-glm(Total.Number.of.Family.members~Total.Household.Income+Total.Food.Expenditure +Household
family = "poisson")
```

Look for outliers in the model

```
library(car)
outlierTest(model.best)
```

```
     rstudent unadjusted p-value Bonferroni p
944 -5.065151         4.0808e-07   0.00050969
```

Remove the row of outliers

```
data<-data[-944,]
model.best<-glm(Total.Number.of.Family.members~Total.Household.Income+Total.Food.Expenditure +Household
family = "poisson")
outlierTest(model.best)
```

```
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
    rstudent unadjusted p-value Bonferroni p
709 -2.89874          0.0037467           NA
```

Without outliers, the best model is obtained

```
summary(model.best)
```

```
Call:
glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + Household.Head.Sex + Household.Head.Age +
    Type.of.Household + House.Age + Number.of.bedrooms, family = "poisson",
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7839  -0.6516  -0.1001   0.4892   2.7201

Coefficients:
                                                    Estimate Std. Error
(Intercept)                                        1.565e+00  7.977e-02
Total.Household.Income                            -5.150e-07  7.839e-08
Total.Food.Expenditure                             6.114e-06  4.521e-07
Household.Head.SexMale                             2.415e-01  3.733e-02
Household.Head.Age                                -5.273e-03  1.089e-03
Type.of.HouseholdSingle Family                    -3.694e-01  3.038e-02
Type.of.HouseholdTwo or More Nonrelated Persons/Members -5.151e-01  2.449e-01
House.Age                                         -2.896e-03  1.152e-03
Number.of.bedrooms                                -2.997e-02  1.575e-02
                                                  z value Pr(>|z|)
(Intercept)                                        19.622  < 2e-16 ***
Total.Household.Income                             -6.570 5.03e-11 ***
Total.Food.Expenditure                             13.524  < 2e-16 ***
Household.Head.SexMale                              6.470 9.82e-11 ***
Household.Head.Age                                 -4.842 1.29e-06 ***
Type.of.HouseholdSingle Family                    -12.159  < 2e-16 ***
Type.of.HouseholdTwo or More Nonrelated Persons/Members  -2.104   0.0354 *
House.Age                                          -2.515   0.0119 *
Number.of.bedrooms                                 -1.902   0.0571 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1373.6  on 1247  degrees of freedom
Residual deviance:  857.2  on 1239  degrees of freedom
AIC: 4900.6

Number of Fisher Scoring iterations: 4
```

Test the goodness of fit of Poisson model

```
library(epiDisplay)
poisgof(model.best)
```

```
$results
```

```
[1] "Goodness-of-fit test for Poisson assumption"

$chisq
[1] 857.1986

$df
[1] 1239

$p.value
[1] 1
```

The p value is 1, which indicates that the goodness of fit of the model is good.

Coefficient and interpretation of model

```
exp(coef(model.best))
```

```
                                            (Intercept) 
                                              4.7835932 
                                 Total.Household.Income 
                                              0.9999995 
                                 Total.Food.Expenditure 
                                              1.0000061 
                                 Household.Head.SexMale 
                                              1.2731562 
                                     Household.Head.Age 
                                              0.9947412 
                         Type.of.HouseholdSingle Family 
                                              0.6911276 
Type.of.HouseholdTwo or More Nonrelated Persons/Members 
                                              0.5974600 
                                              House.Age 
                                              0.9971080 
                                     Number.of.bedrooms 
                                              0.9704752 
```

In the MIMAROPA region, all variables except Number of. Both bedrooms and Electricity show significance. While keeping other variables unchanged, the annual household income (in Philippines Peso) will be increased by 1 unit, and the number of people living in the house will be multiplied by 0.999995. Annual expenditure by the household on food changes, the number of epilepsy will be multiplied by 1.0000061. If the gender of head of the houses sex is male, the number of people living in the house will be multiplied by 1.2731562, indicating that the owner is male, which has a positive impact on the increase of the number of people living in the room. The number of people living in the house will be multiplied by 0.9947412 for each additional year of head of the houses age. In the relationship between the group of people living in the house, both single family and two or more nonrelated persons / members will both have a negative impact on the increase of the number of people living in the room. The number of people living in the house will be multiplied by 0.9971080 for each year of age of the building.

Conclusions and Future Work