Group2

Anwen Jin

2023-03-14

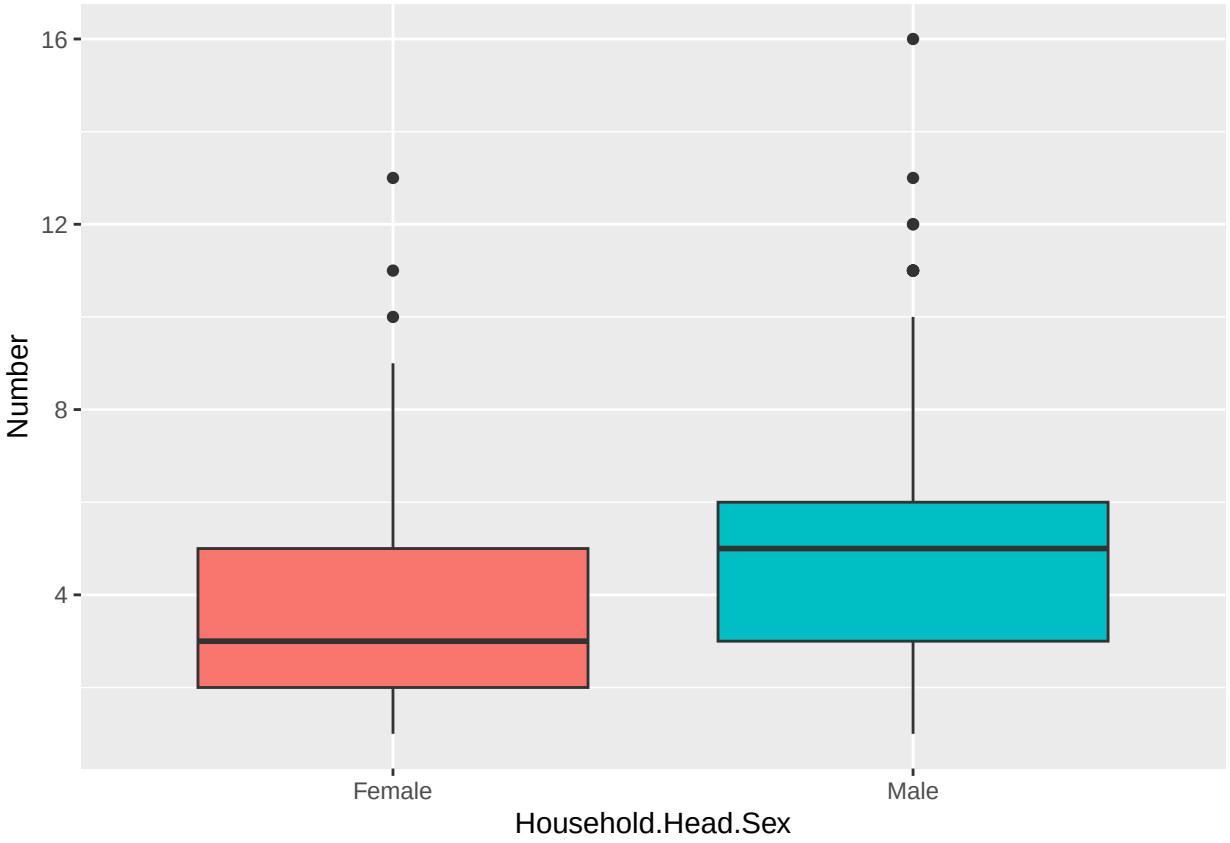
Dataset

The dataset comes from the FIES (Family Income and Expenditure Survey) recorded in the Philippines. The survey, which is undertaken every three years, is aimed at providing data on family income and expenditure. The datasets obtained from this survey is from the region,IVB - MIMAROPA, across the Philippines and the following variables, recorded by household.

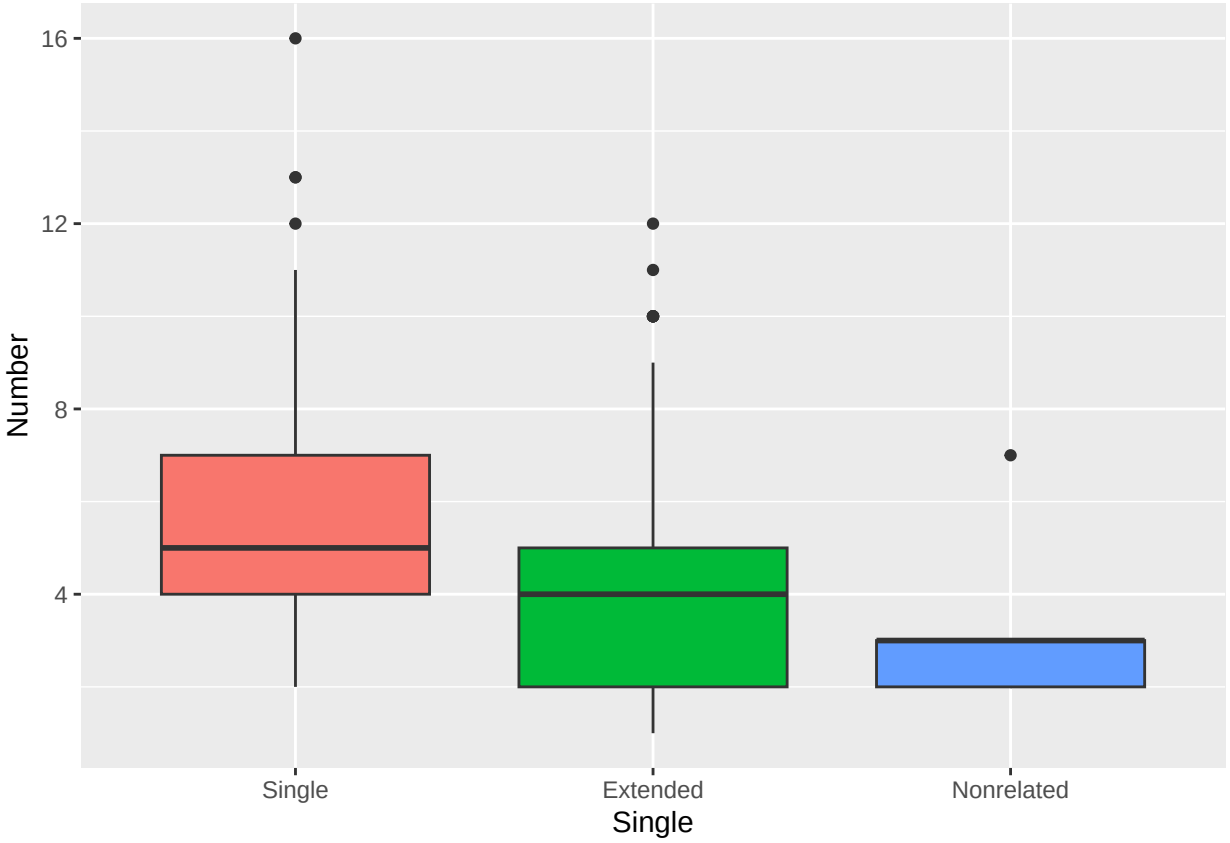
Variable analysis

Because the response variable,Number of people living in the house, is integer data. In that case, GLM model for Poisson distribution is picked to fit the data.

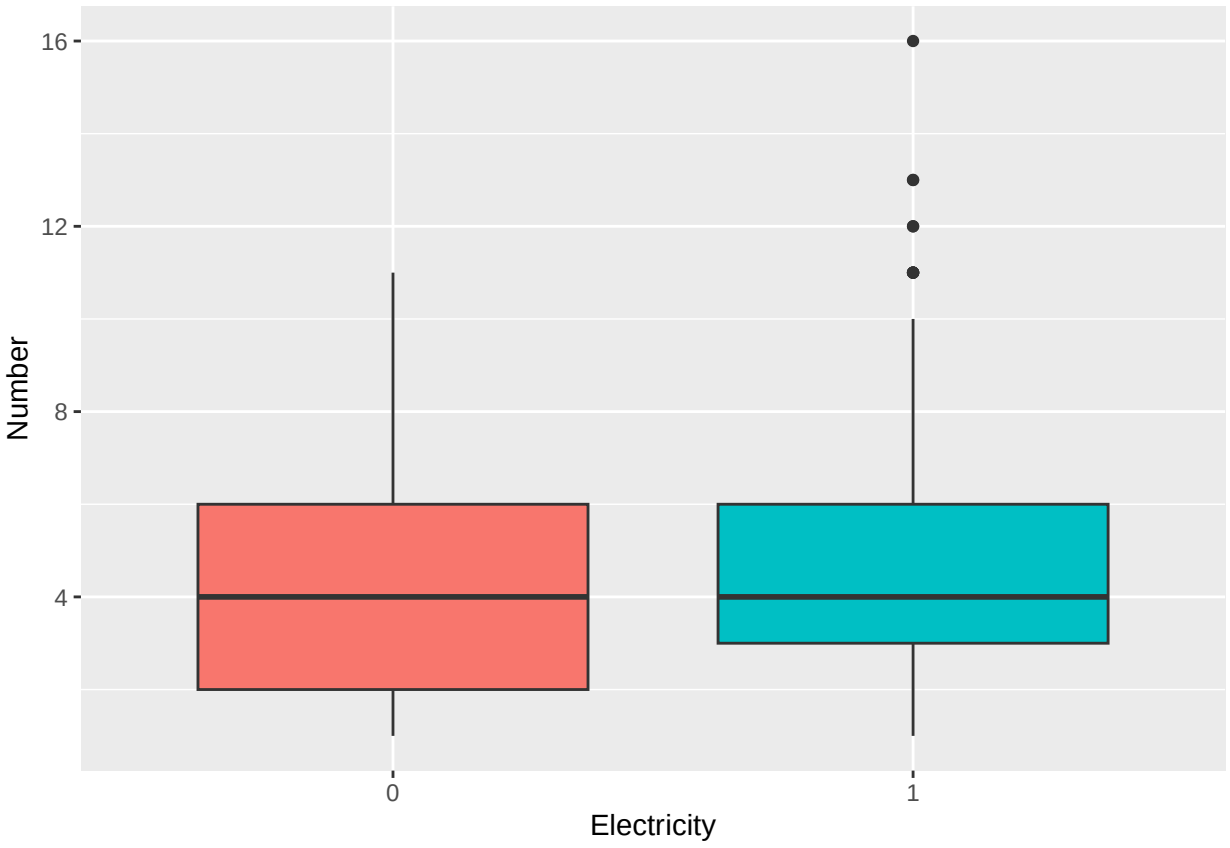
Relationship between variables



As can be seen from the p2, there are more people living in the house when the owner is a man.



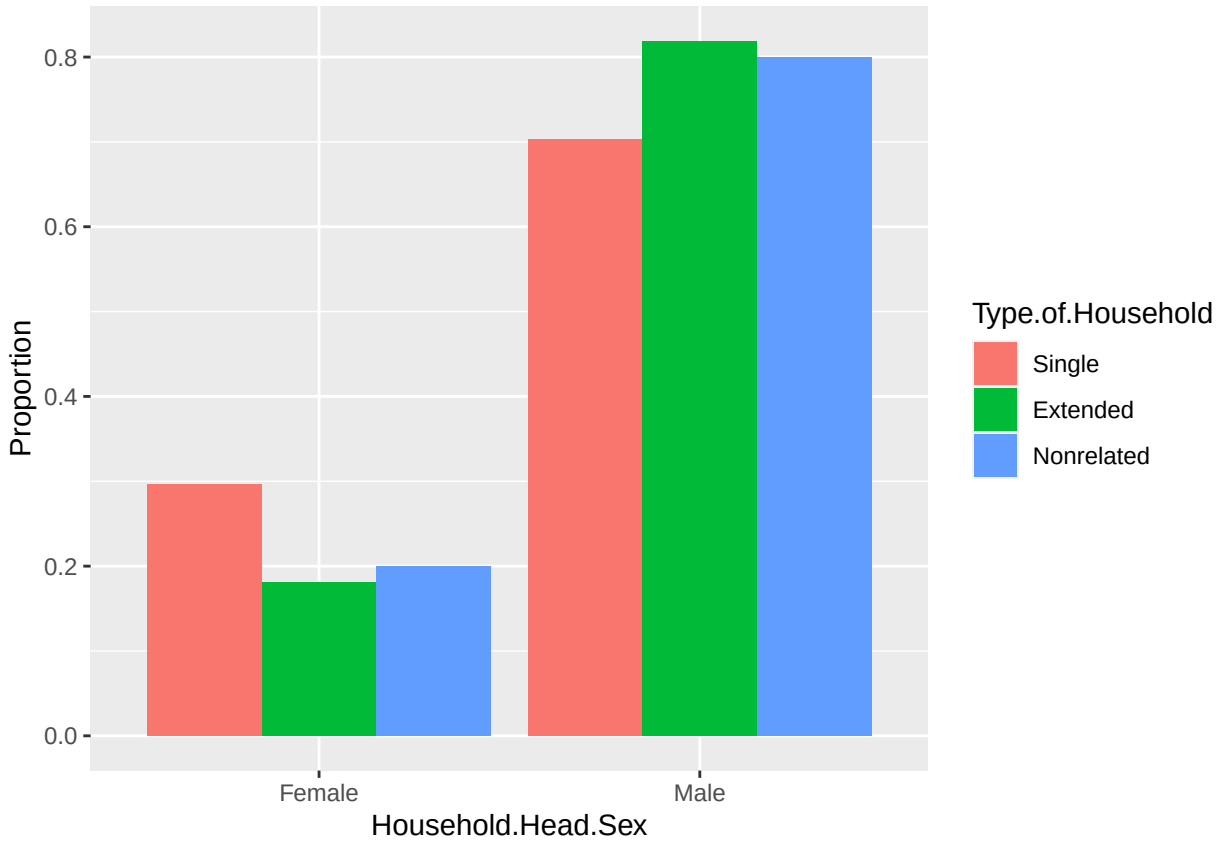
It can be seen from p3 that there are more single people



It can be seen from p4 that the number of household members with or without electricity is similar, and they have the same median.

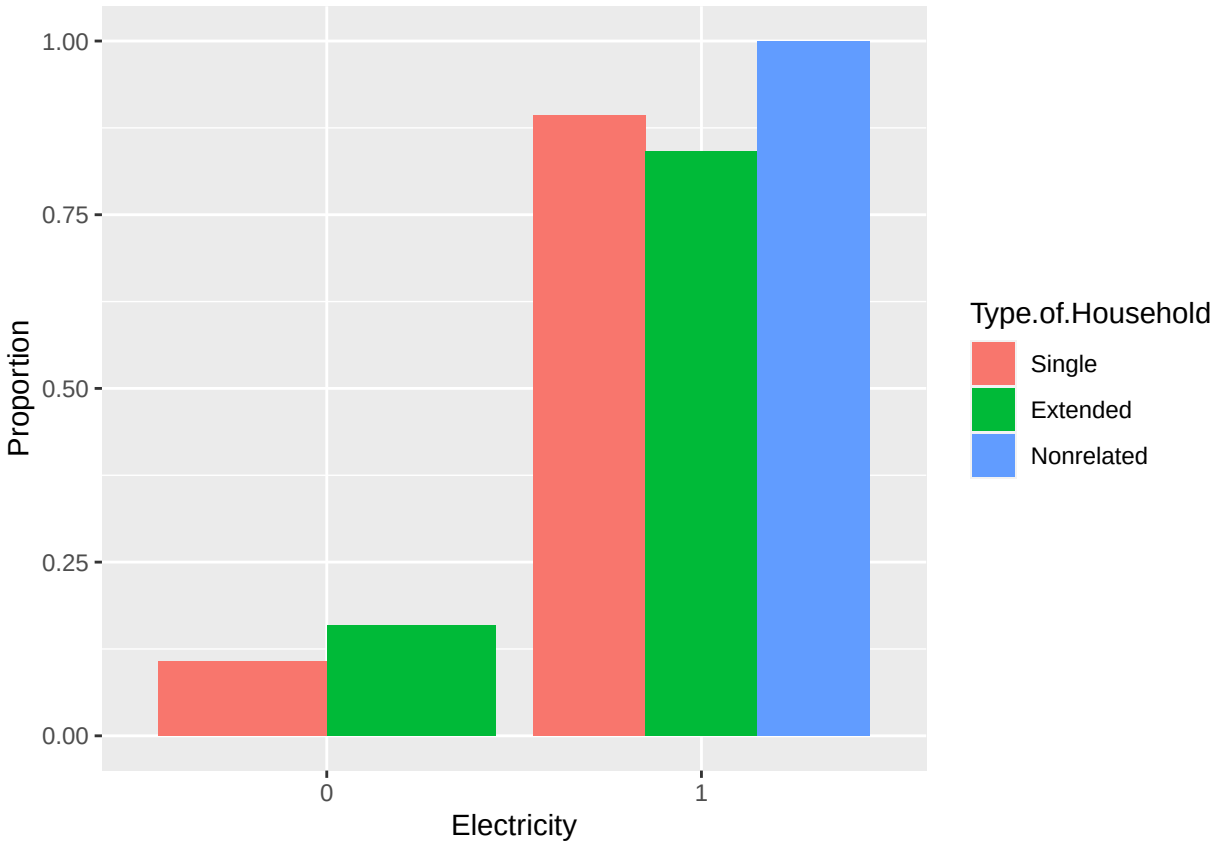
```
##  Type.of.Household      Female        Male
##             Single 29.7% (102) 70.3% (242)
##           Extended 18.1% (163) 81.9% (737)
##         Nonrelated 20.0%   (1) 80.0%   (4)
```

```
## Warning: The dot-dot notation (‘..prop..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(prop)‘ instead.
```



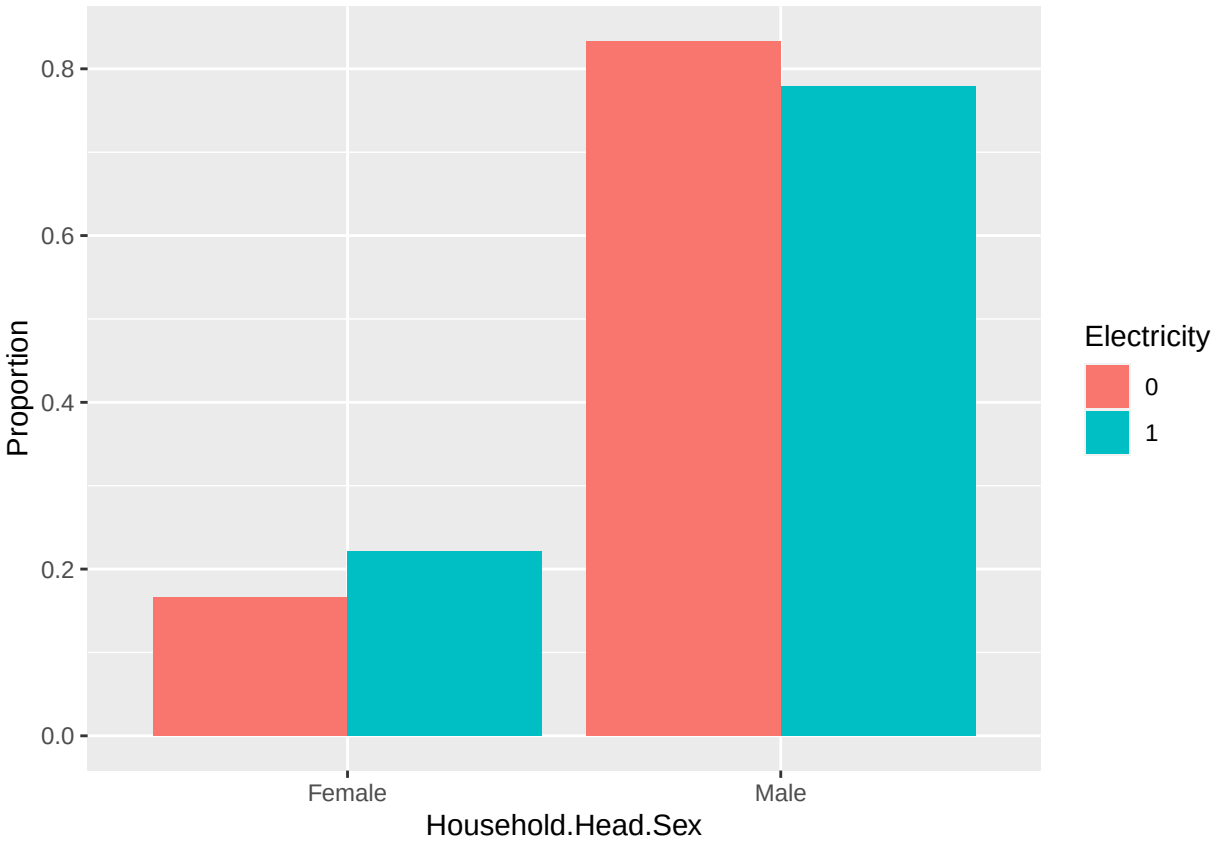
From p5, it can be seen that when the house is owned by the man, it is mostly a extended family. Women were more likely to live alone.

```
##  Type.of.Household           0            1
##             Single 10.8%  (37)  89.2% (307)
##           Extended 15.9% (143)  84.1% (757)
##         Nonrelated  0.0%   (0) 100.0%   (5)
```



p6:When the house has no electricity, more people choose to live alone. When electricity is available, the proportion of extended families is higher.

```
##  Electricity      Female        Male
##            0 16.7%  (30) 83.3% (150)
##            1 22.1% (236) 77.9% (833)
```



p7:When the house is owned by the women, it is mostly a family with electricity. For men, the proportion of electricity use is lower.

##Models

```
##
## Call:
## glm(formula = Total.Number.of.Family.members ~ Number.of.bedrooms +
##     Total.Household.Income + Total.Food.Expenditure + Household.Head.Age +
##     House.Floor.Area + House.Age + Electricity + Household.Head.Sex +
##     Type.of.Household, 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 z value Pr(>|z|)
## (Intercept)                  1.671e+00  8.230e-02  20.299  < 2e-16 ***
## Number.of.bedrooms          -2.366e-02  1.680e-02  -1.409   0.1589
## Total.Household.Income      -4.266e-07  7.596e-08  -5.616 1.96e-08 ***
## Total.Food.Expenditure       5.239e-06  4.066e-07  12.886  < 2e-16 ***
## Household.Head.Age          -5.818e-03  1.080e-03  -5.386 7.21e-08 ***
## House.Floor.Area            -9.056e-05  3.033e-04  -0.299   0.7653
## House.Age                   -2.451e-03  1.177e-03  -2.082   0.0374 *
## Electricity1                -5.232e-02  4.048e-02  -1.293   0.1961
## Household.Head.SexMale       2.418e-01  3.739e-02   6.467 1.00e-10 ***
## Type.of.HouseholdExtended   -3.732e-01  3.047e-02 -12.250  < 2e-16 ***
```

```
## Type.of.HouseholdNonrelated -5.036e-01  2.447e-01  -2.058   0.0396 *
## ---
## 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
```

Notice that the coefficient of number of bedrooms is negative,suggesting that Families with more bedrooms do not necessarily have many family members.However, the coefficient of age is not significant (p-value of 0.1589).Again, others coefficient are negative except for the food expenditure and household hear sexmale.We select data with p-values less than 0.05 for analysis.

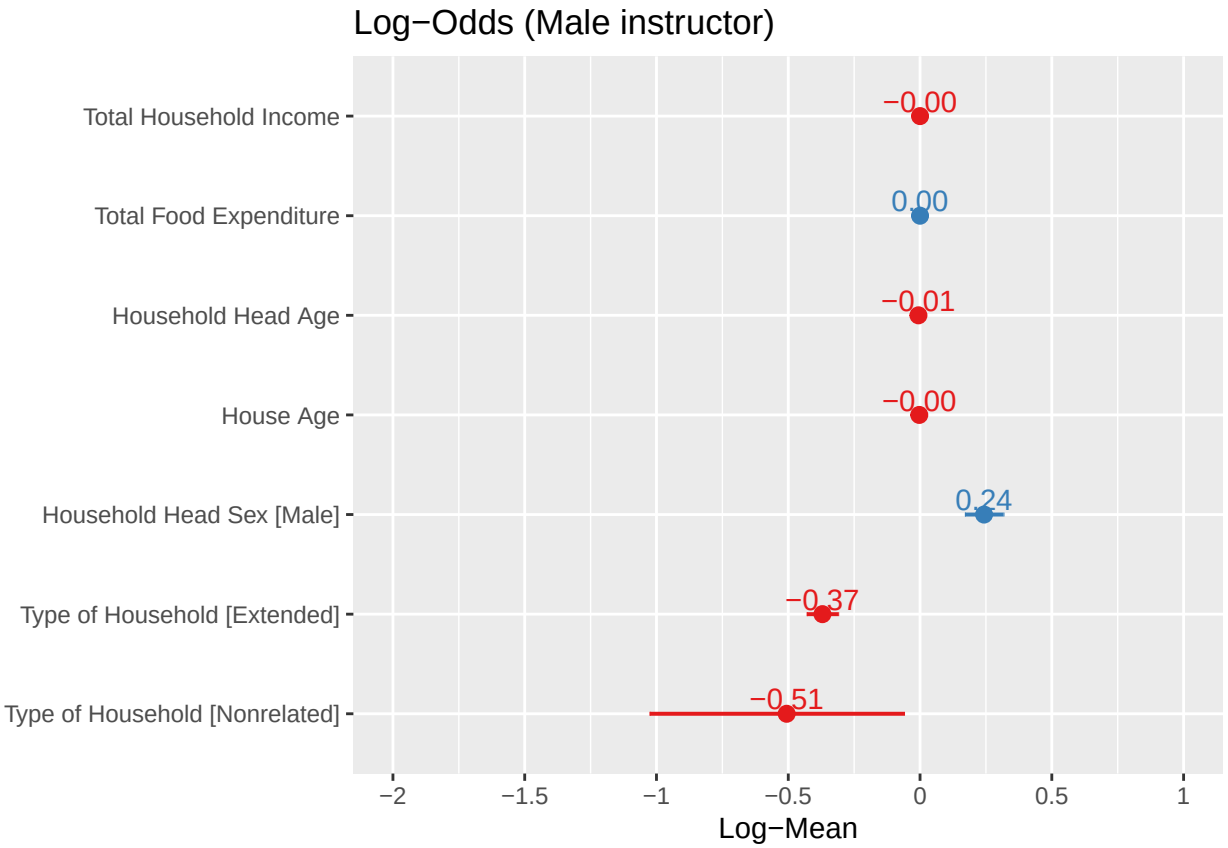
from the p-value, whose <0.05

```
##
## Call:
## glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
##     Total.Food.Expenditure + Household.Head.Age + House.Age +
##     Household.Head.Sex + Type.of.Household, family = poisson(),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.5525  -0.6588  -0.1129   0.5048   2.7146
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.617e+00  7.735e-02  20.904  < 2e-16 ***
## Total.Household.Income      -4.579e-07  7.492e-08  -6.112 9.86e-10 ***
## Total.Food.Expenditure       5.068e-06  4.008e-07  12.644  < 2e-16 ***
## Household.Head.Age          -6.089e-03  1.069e-03  -5.695 1.23e-08 ***
## House.Age                   -2.915e-03  1.150e-03  -2.535   0.0113 *
## Household.Head.SexMale       2.430e-01  3.733e-02   6.509 7.58e-11 ***
## Type.of.HouseholdExtended   -3.704e-01  3.038e-02 -12.190  < 2e-16 ***
## Type.of.HouseholdNonrelated -5.061e-01  2.446e-01  -2.069   0.0386 *
## ---
## 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:  886.06  on 1241  degrees of freedom
## AIC: 4930.9
##
## Number of Fisher Scoring iterations: 4
```

We see the coefficient for food expenditure and household head sexmale are positive, indicating Households with higher total food expenditures have more members,and if the head of the household is male, he has more family members.The p-value of all parameters is significant(<0.05), So they have a significant effect on the number of family members.

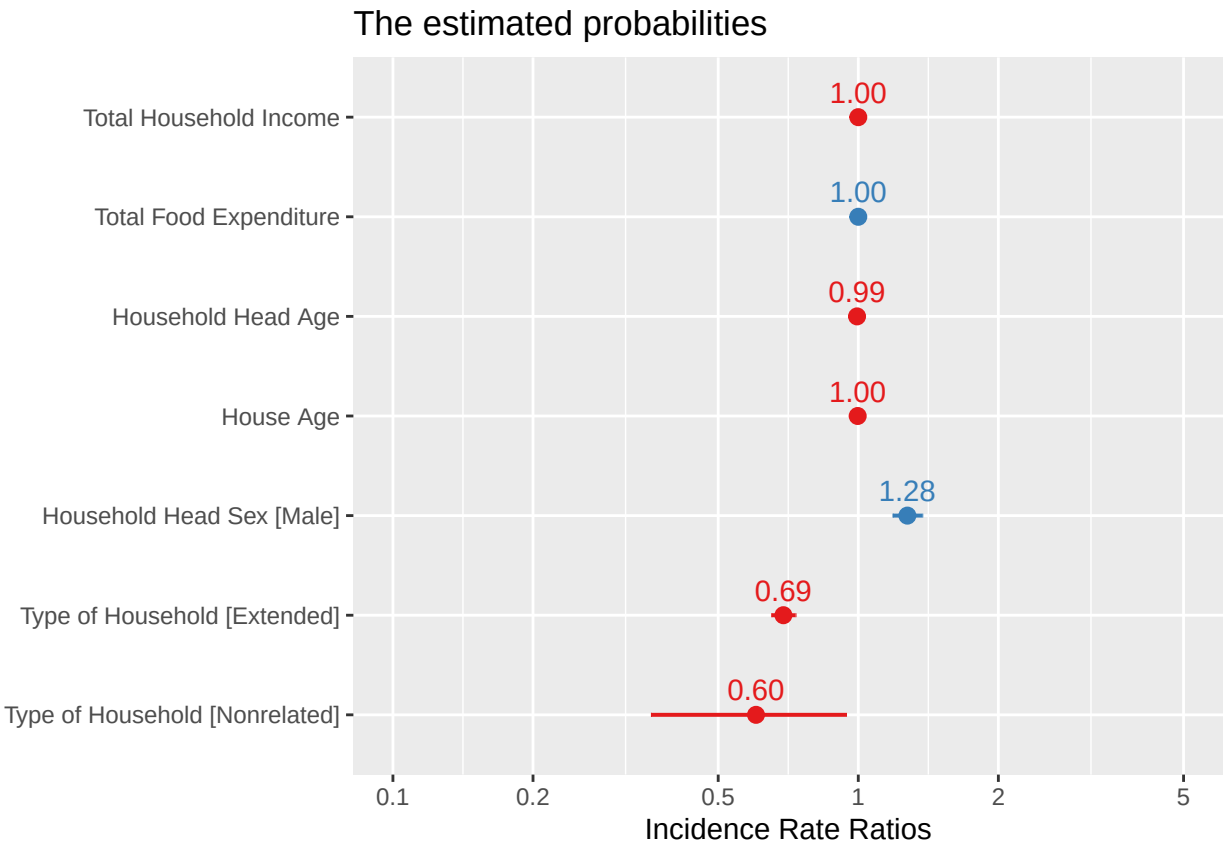
```
## Waiting for profiling to be done...
```

	2.5 %	97.5 %
(Intercept)	1.4648617	1.7680597
Total.Household.Income	-0.0000006	-0.0000003
Total.Food.Expenditure	0.0000043	0.0000058
Household.Head.Age	-0.0081841	-0.0039935
House.Age	-0.0051863	-0.0006778
Household.Head.SexMale	0.1703110	0.3166476
Type.of.HouseholdExtended	-0.4297918	-0.3106874
Type.of.HouseholdNonrelated	-1.0258509	-0.0608476



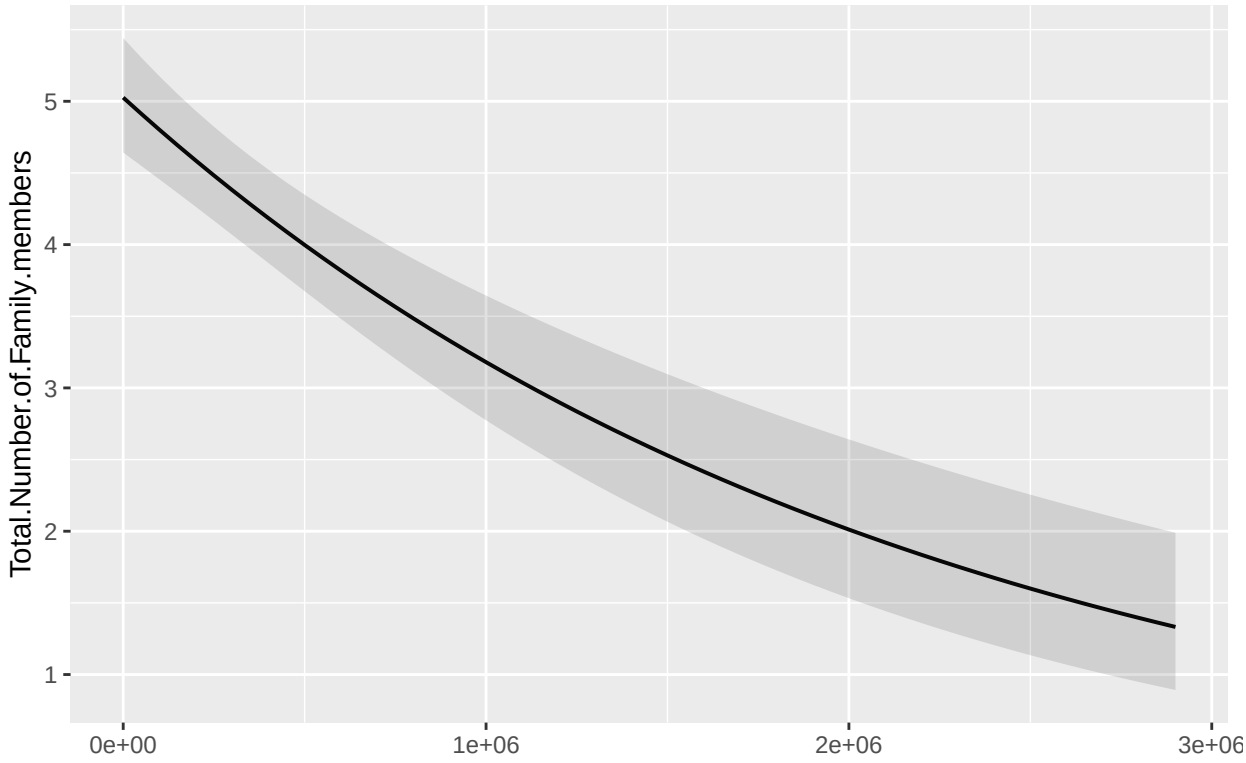
We interpret the odds ratios as follows: household head sex(male) were 0.24 times those of female.

```
##                 (Intercept)      Total.Household.Income
##                   5.0374649                   0.9999995
##      Total.Food.Expenditure          Household.Head.Age
##                   1.0000051                   0.9939299
##                   House.Age      Household.Head.SexMale
##                   0.9970889                   1.2750109
##   Type.of.HouseholdExtended Type.of.HouseholdNonrelated
##                   0.6904735                   0.6028550
```

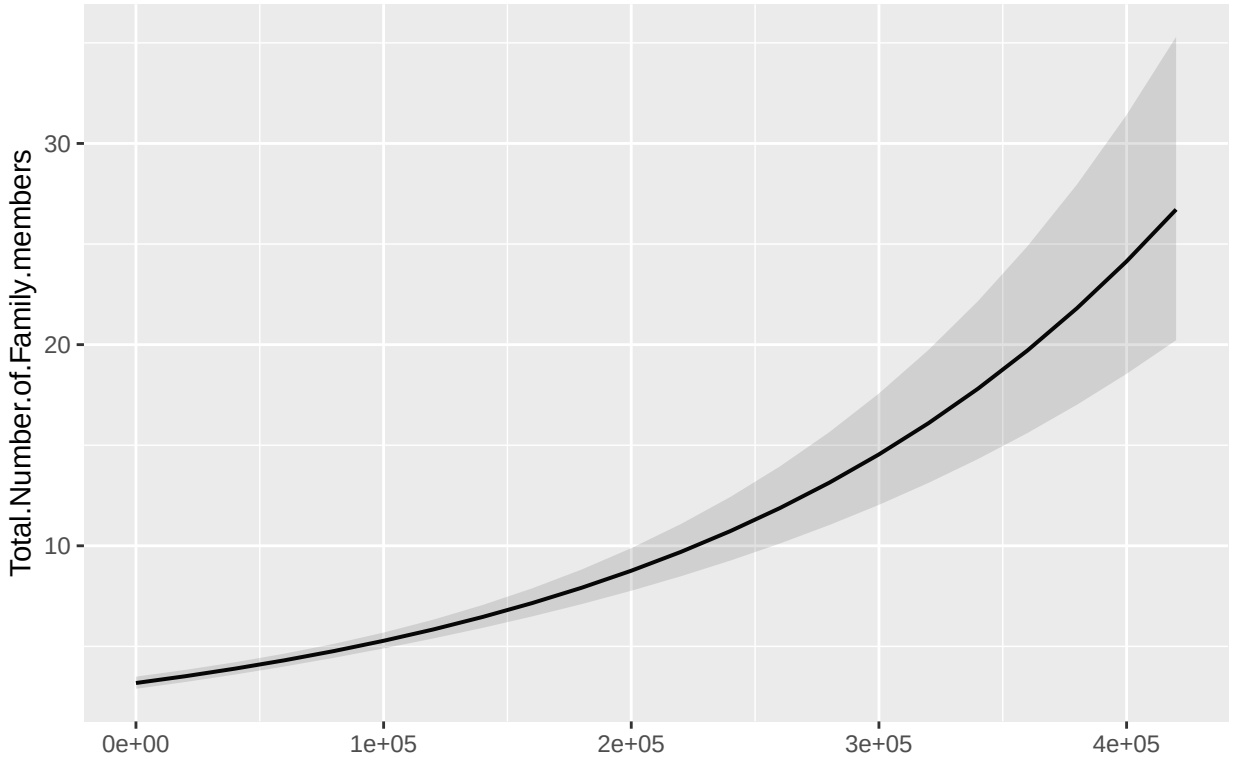


We interpret the odds ratios as follows: household head sex(male) were 1.28 times those of female, extended family were 0.69 times those of single family.Finally, for extended family increase ,the family members decrease(by a factor of 0.69).

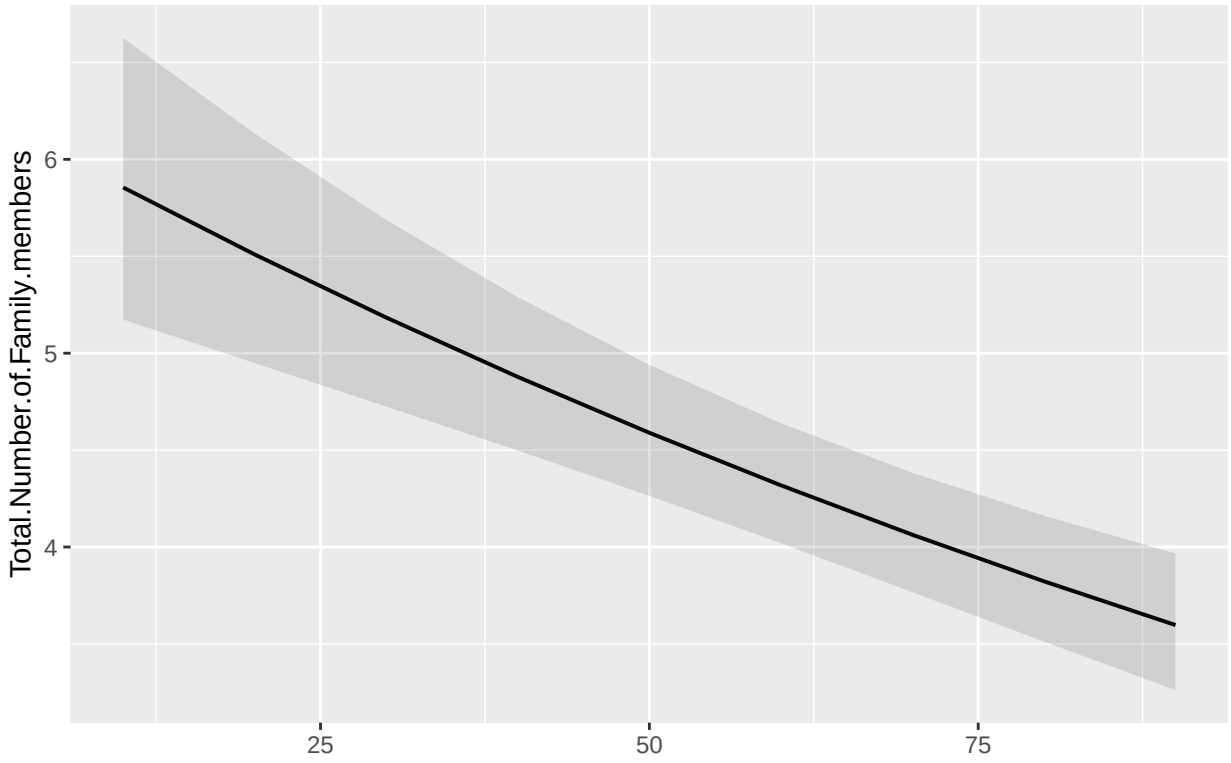
```
## $Total.Household.Income
```



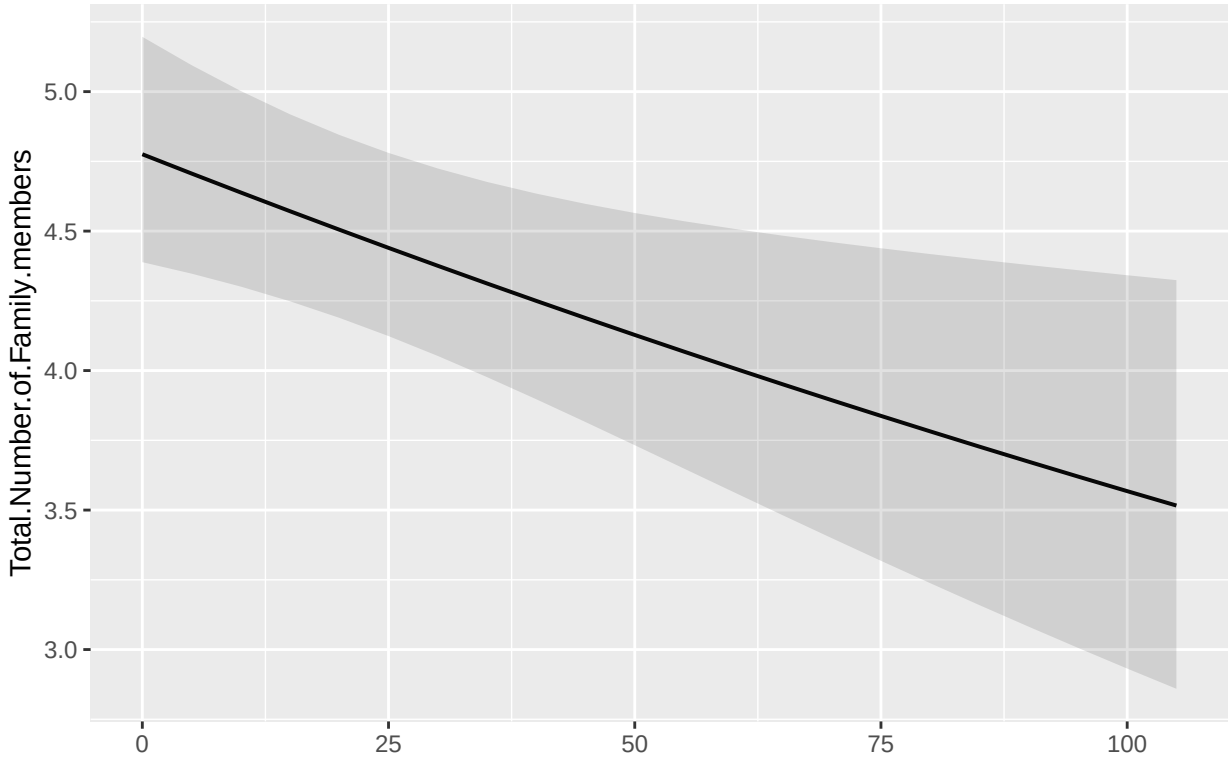

```
##
## $Total.Food.Expenditure
```




```
##
## $Household.Head.Age
```




```
##
## $House.Age
```




```
##
## $Household.Head.Sex
```


```
##
## $Type.of.Household
```