project2_test

Group_01

2021/7/7

	Total.Number.of.Family.members
Total.Number.of.Family.members	1.00000000
Total.Household.Income	0.19228742
Total.Food.Expenditure	0.46924215
Household.Head.Age	-0.06541636
House.Floor.Area	-0.01415702
House.Age	-0.07003586
Number.of.bedrooms	0.07207630
Electricity	0.09193871

	Total.Household.Income	Total.Food.Expenditure
Total.Number.of.Family.members	0.19228742	0.469242145
Total.Household.Income	1.00000000	0.611494530
Total.Food.Expenditure	0.61149453	1.000000000
Household.Head.Age	0.06280405	-0.051724735
House.Floor.Area	0.23413840	0.124320633
House.Age	0.02471720	0.006725185
Number.of.bedrooms	0.44137375	0.355734454
Electricity	0.14866655	0.198610366

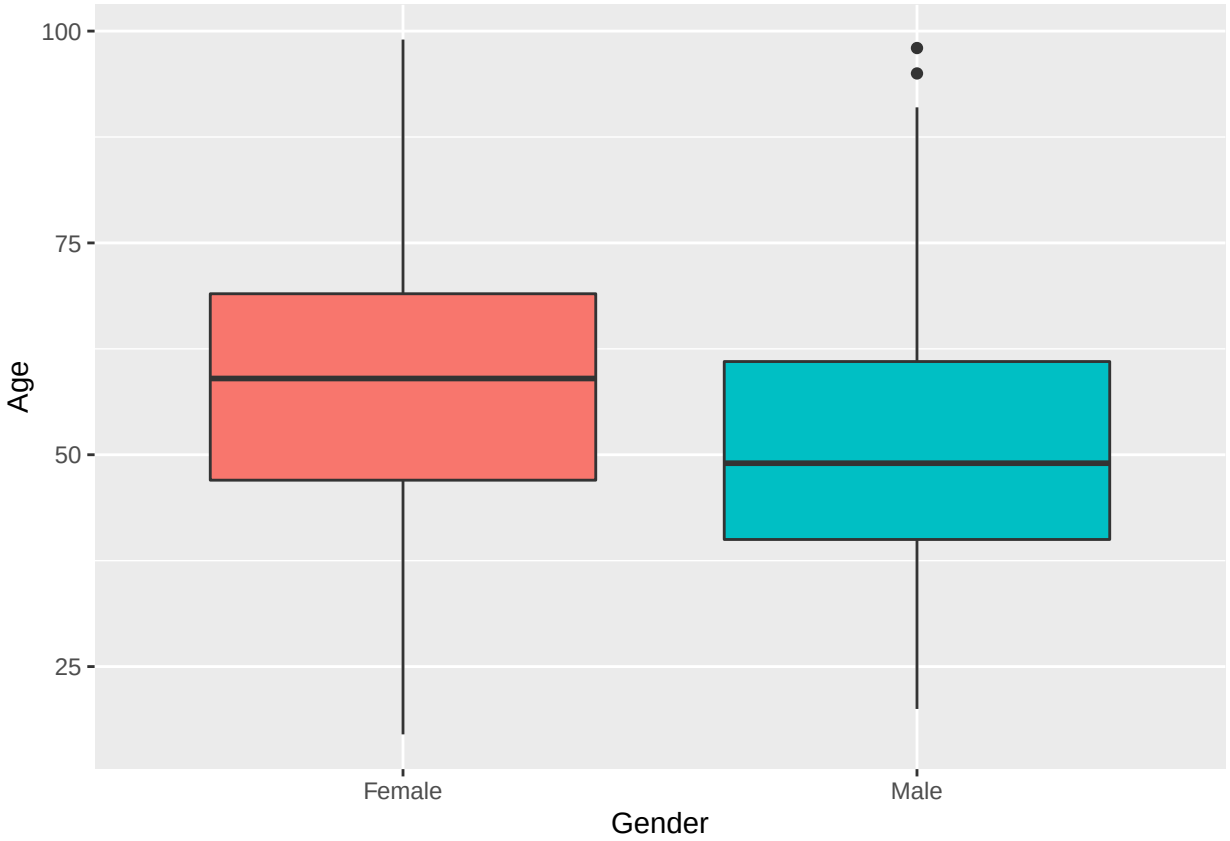
	Household.Head.Age	House.Floor.Area	House.Age
Total.Number.of.Family.members	-0.06541636	-0.01415702	-0.070035856
Total.Household.Income	0.06280405	0.23413840	0.024717197
Total.Food.Expenditure	-0.05172474	0.12432063	0.006725185
Household.Head.Age	1.00000000	0.09057216	0.218079293
House.Floor.Area	0.09057216	1.00000000	0.074265080
House.Age	0.21807929	0.07426508	1.000000000
Number.of.bedrooms	0.15415511	0.37399081	0.123180471
Electricity	-0.01304412	0.10693465	0.085327324

	Number.of.bedrooms	Electricity
Total.Number.of.Family.members	0.0720763	0.09193871
Total.Household.Income	0.4413738	0.14866655
Total.Food.Expenditure	0.3557345	0.19861037
Household.Head.Age	0.1541551	-0.01304412
House.Floor.Area	0.3739908	0.10693465
House.Age	0.1231805	0.08532732
Number.of.bedrooms	1.0000000	0.21376315
Electricity	0.2137632	1.00000000

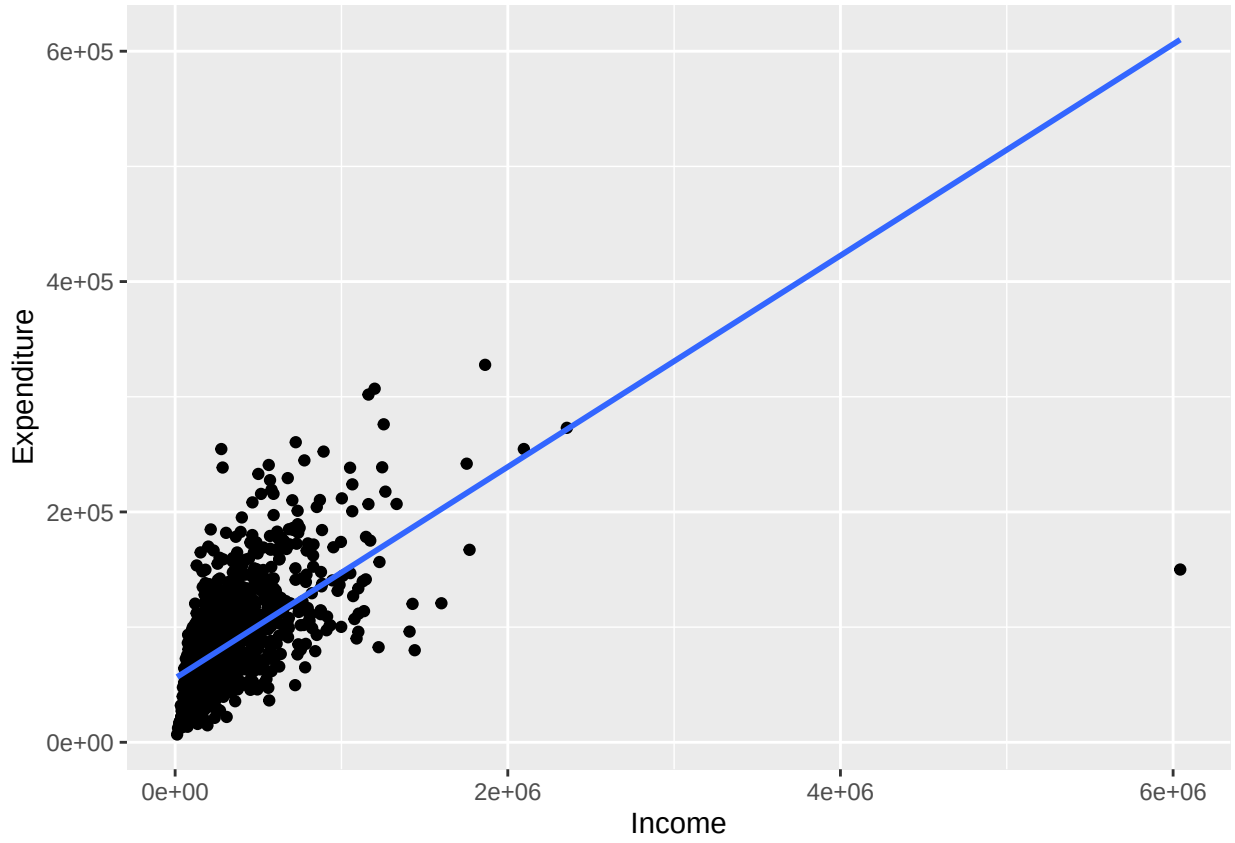
Table 1: Summary statistics

Variable	Missing	Complete	Mean	SD	Min.	1st Q.	Median	3rd Q.	
Total.Number.of.Family.members	0	1	4.67	2.33	1	3	4	6	
Total.Household.Income	0	1	269540.48	274564.17	11988	118565	188580	328335	6
Total.Food.Expenditure	0	1	80352.78	41194.36	6781	51922	73578	98493	
Household.Head.Age	0	1	52.23	14.52	17	41	52	63	
House.Floor.Area	0	1	90.92	99.20	5	32	54	102	
House.Age	0	1	22.98	15.32	0	12	20	31	
Number.of.bedrooms	0	1	2.26	1.44	0	1	2	3	
Electricity	0	1	0.93	0.26	0	1	1	1	

Gender&age



balance

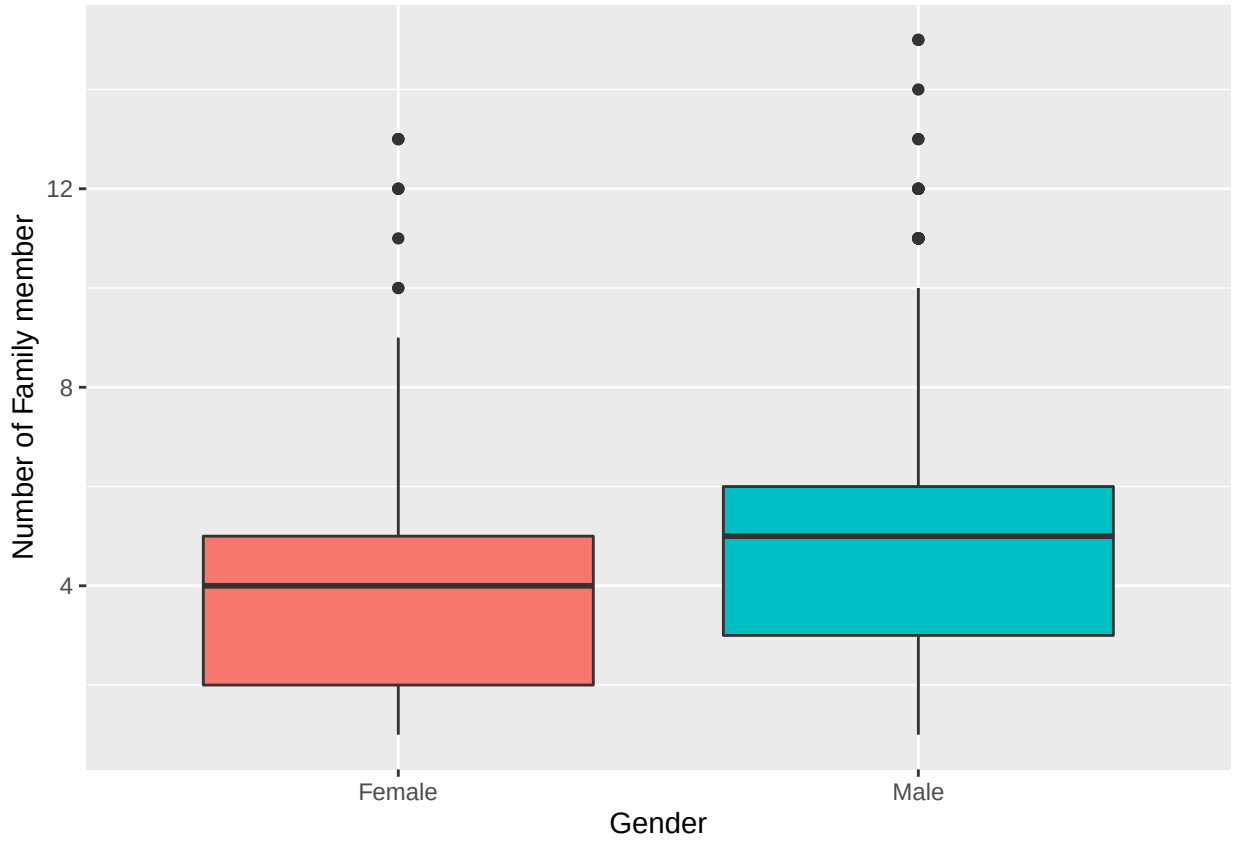


Model

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

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.5671  -1.4626  -0.3084   1.2037  10.7417

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             2.746e+00  2.667e-01  10.298  < 2e-16 ***
Total.Household.Income -1.022e-06  2.384e-07  -4.287 1.91e-05 ***
Total.Food.Expenditure  3.197e-05  1.540e-06  20.759  < 2e-16 ***
Household.Head.Age     -4.491e-04  3.520e-03  -0.128  0.89850
House.Floor.Area       -7.261e-04  5.357e-04  -1.355  0.17550
House.Age              -9.472e-03  3.301e-03  -2.870  0.00416 **
Number.of.bedrooms     -9.756e-02  4.121e-02  -2.367  0.01802 *
Electricity             1.696e-01  1.929e-01   0.879  0.37955
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for gaussian family taken to be 4.130968)

    Null deviance: 9384.0  on 1724  degrees of freedom
Residual deviance: 7092.9  on 1717  degrees of freedom
AIC: 7352.3

Number of Fisher Scoring iterations: 2


Call:
glm(formula = Total.Number.of.Family.members ~ Total.Household.Income +
    Total.Food.Expenditure + House.Age + Number.of.bedrooms,
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.5796  -1.4561  -0.3048   1.1778  10.6187

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             2.828e+00  1.375e-01  20.560  < 2e-16 ***
Total.Household.Income -1.061e-06  2.364e-07  -4.487 7.71e-06 ***
Total.Food.Expenditure  3.229e-05  1.513e-06  21.340  < 2e-16 ***
House.Age              -9.507e-03  3.223e-03  -2.950  0.00322 **
Number.of.bedrooms     -1.103e-01  3.855e-02  -2.862  0.00425 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4.129948)

    Null deviance: 9384.0  on 1724  degrees of freedom
Residual deviance: 7103.5  on 1720  degrees of freedom
AIC: 7348.8

Number of Fisher Scoring iterations: 2
```

Family number & Gender



Log-odds

```
Call:
glm(formula = Household.Head.Sex ~ Total.Number.of.Family.members,
    family = binomial(link = "logit"), data = data.sex_number)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4219   0.4705   0.6602   0.7163   0.9054

Coefficients:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.49674    0.13174   3.771 0.000163 ***
Total.Number.of.Family.members  0.18319    0.02844   6.442 1.18e-10 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1790.9  on 1724  degrees of freedom
Residual deviance: 1745.4  on 1723  degrees of freedom
AIC: 1749.4
```

```
Number of Fisher Scoring iterations: 4
```

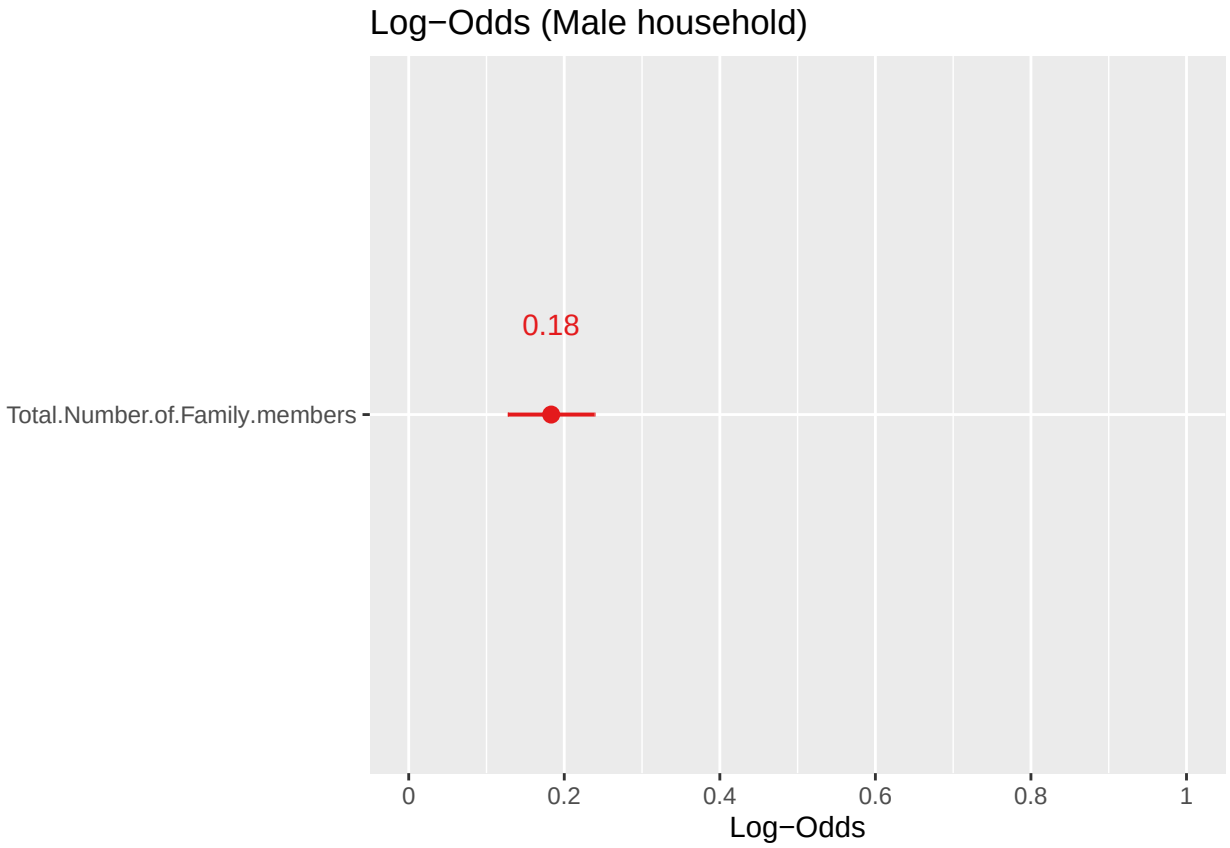
```
[1] "Female" "Male"
```

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot \text{number of family members} = 0.5 + 0.18 \cdot \text{number of family members},$$

Where $p = Prob(Male)$ and $1 - p = Prob(Female)$. Hence, the log-odds of the household being male increase by 0.18 for every one unit increase in number of family members. This provides us with a point estimate of how the log-odds changes with age.

However, we are also interested in producing a 95% confidence interval for these log-odds.

	2.5 %	97.5 %
(Intercept)	0.2388990	0.7555347
Total.Number.of.Family.members	0.1282353	0.2397474



Now, let's add the estimates of the log-odds to our data set: