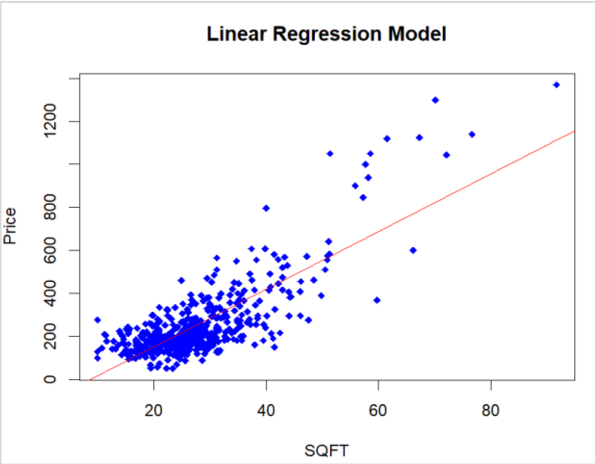
2.17



(a)(b)

PRICE = -115.4236+13.4029xSQFT

Holding other conditions constant, for every additional 100 square feet of house, the expected house price will increase by $13402.9.

The estimated intercept is -115423.6, meaning that a house with an area of 0 square feet is expected to have a price of $ -115423.6

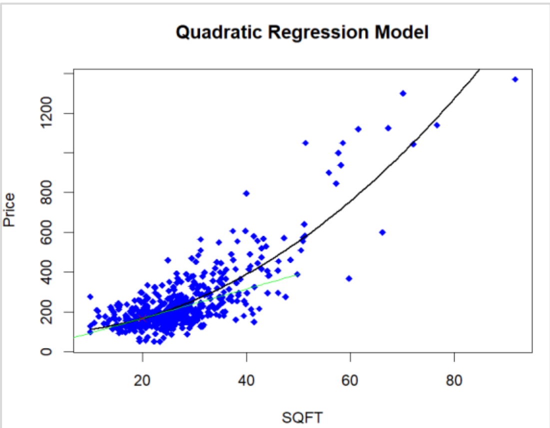
(c)

PRICE = 93.5659+0.1845 x SQFT^2

Marginal effect = 2 x 0.1845 x 20 = 7.3807

When the house size is 2,000 square feet, adding an additional 100 square feet of space is expected to increase the house price by $7380.8.

(d)



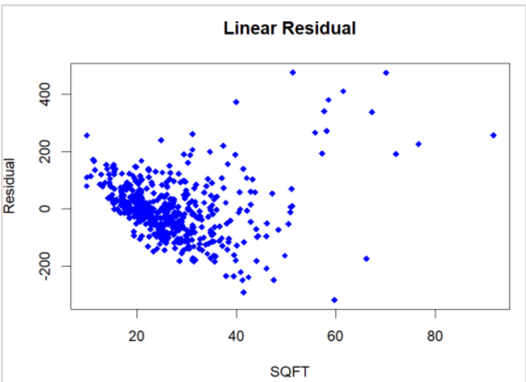
(e)Elasticity = 0.8819511

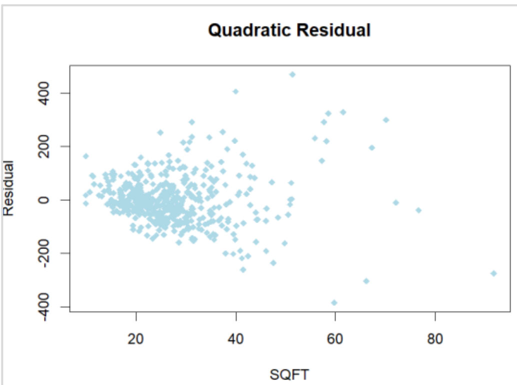
(g)

Because the SSE of quadratic regression model is smaller than the one of the linear regression model, the quadratic regression model is the “better-fitting” model.

```
> SSE_lin
[1] 5262847
> SSE_quad
[1] 4222356
```

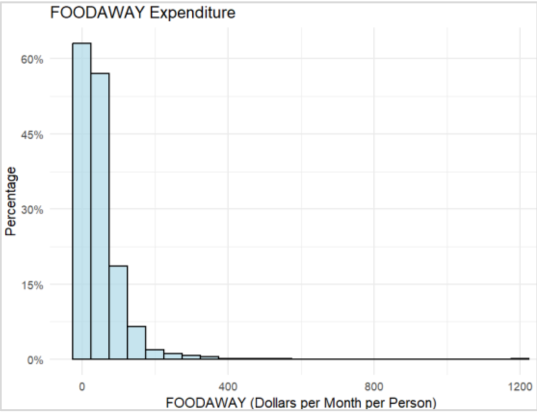
(f) The residual pattern does not exhibit a random distribution. As SQFT increases, the variability of the residuals also increases, indicating a potential violation of the homoskedasticity assumption.





2.25

(a)



Mean = 49.27

Median = 32.55,

25th percentiles = 12.04

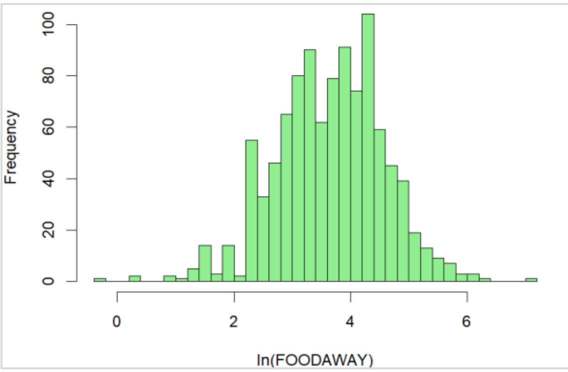
75th percentiles = 67.5025

```
> summary(foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   12.04   32.55   49.27   67.50 1179.00
```

(b)

	Advanced Degree	College Degree	No degree
N	257	369	574
Mean	73.15494	48.59718	39.01017
Median	48.15	36.11	26.02

(c) Because the log() function cannot handle 0 or negative values, ln(FOODAWAY) will have fewer valid observations.




```
> summary(foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   12.04   32.55   49.27   67.50 1179.00
```

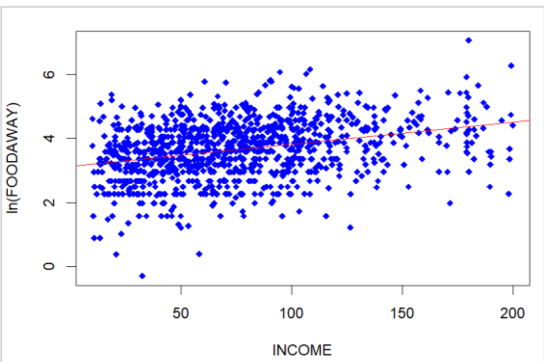
```
Observations of FOODAWAY: 1200
Observations of ln(FOODAWAY):   1022
```

(d)

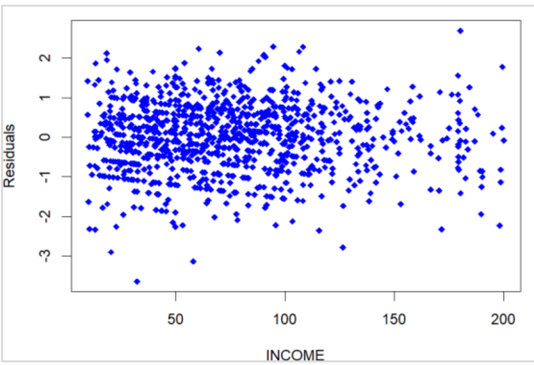
$\ln(FOODAWAY) = 3.1293004 + 0.0069017\ INCOME$

Holding other conditions constant, we estimate that each additional \$100 household income increases food away expenditures per person of about 0.69%.

(e)It shows a positive association between ln(FOODAWAY) and INCOME.

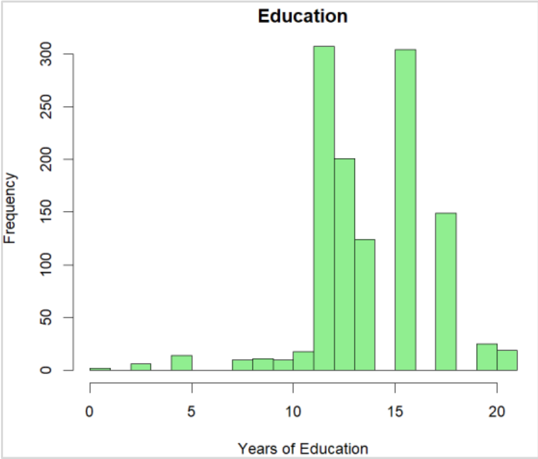


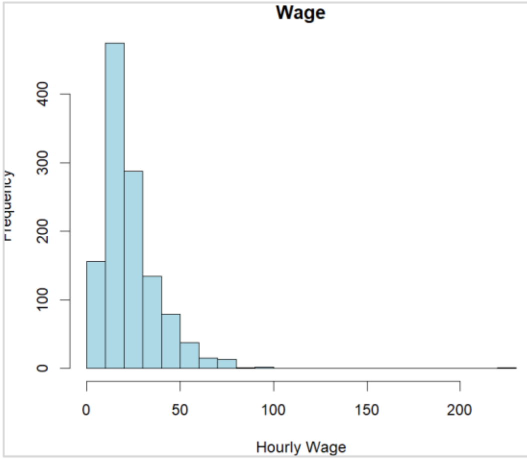
(f)The residuals looks like randomly distributed with no obvious patterns. There are few observations at higher incomes.



2.28

(a)






```
> summary(cps5_small$wage)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.94   13.00   19.30   23.64   29.80  221.10
> summary(cps5_small$educ)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    12.0    14.0    14.2    16.0    21.0
```

The EDUC histogram might show clustering around certain education levels.

The WAGE histogram might show right skewness.

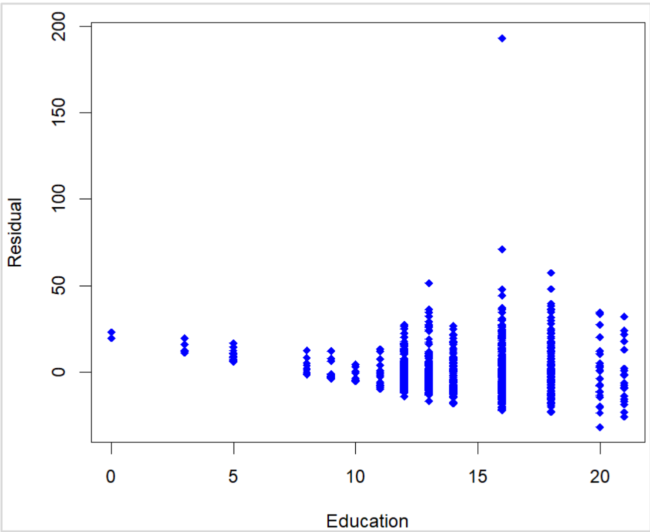
(b)

WAGE = -10.4 + 2.3968 x EDUC

The intercept is -10.4, means when the year of education is 0, the predicted wage is -10.4.

The slope is 2.3968, means for each additional year of education, the predicted hourly wage increases by $2.40.

(c)



The residual pattern does not exhibit a random distribution. As the year of education increases, the variability of the residuals also increases, indicating a potential violation of the homoskedasticity assumption.

If the OLS assumptions (SR1-SR5) hold, the residuals should be evenly distributed across different levels of EDUC.

(d)

Model of male = -8.2849 + 2.3758 x EDUC

Model of female = -16.6028 + 2.6595 x EDUC

Model of blacks = -6.2541 + 1.9233 x EDUC

Model of whites = -10.475 + 2.418 x EDUC

The slope is slightly higher than for males (2.6595 vs. 2.3758), suggesting that education benefits females slightly more in terms of wage increases.

The return to education for White individuals ($2.42 per hour) is higher than for Black individuals ($1.92 per hour), suggesting racial disparities in wage returns to education.

Females have the highest return to education ($2.66 per year).

Blacks have the lowest return ($1.92 per year).

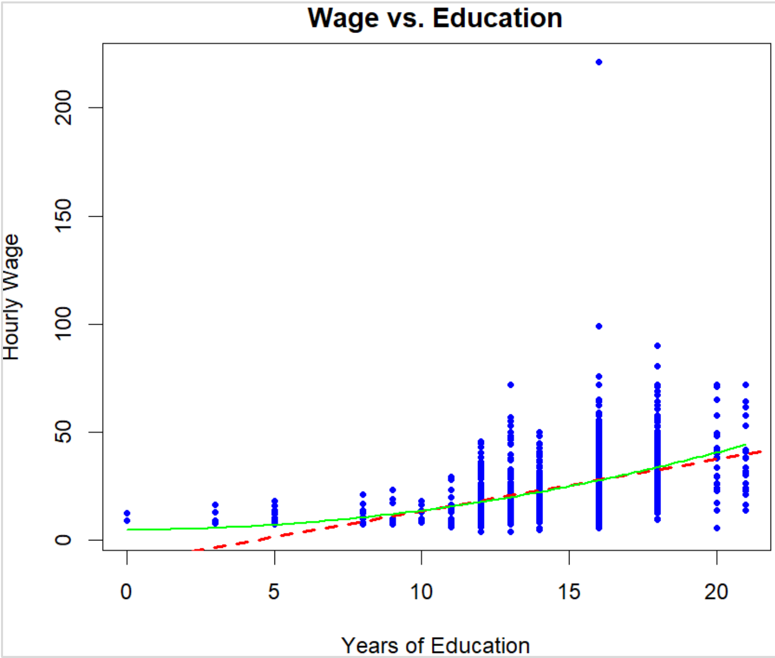
(e)

WAGE = 4.916477 + 0.089134 x $EDUC^2$

The marginal effect is greatest when the years of education are 16.

```
> cat("Marginal effect of EDUC = 12 :", marginal_effect_12, "\n")
Marginal effect of EDUC = 12 : 2.139216
> cat("Marginal effect of EDUC = 16 :", marginal_effect_16, "\n")
Marginal effect of EDUC = 16 : 2.852288
> cat("Marginal effect of linear regression :", linear_effect, "\n")
Marginal effect of linear regression : 2.396761
```

(f)



Red line : Linear model

Green line : Quadratic model

The quadratic model predicts more accurately than the linear model, and the green line is also closer to the actual values when EDUC < 10.