Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

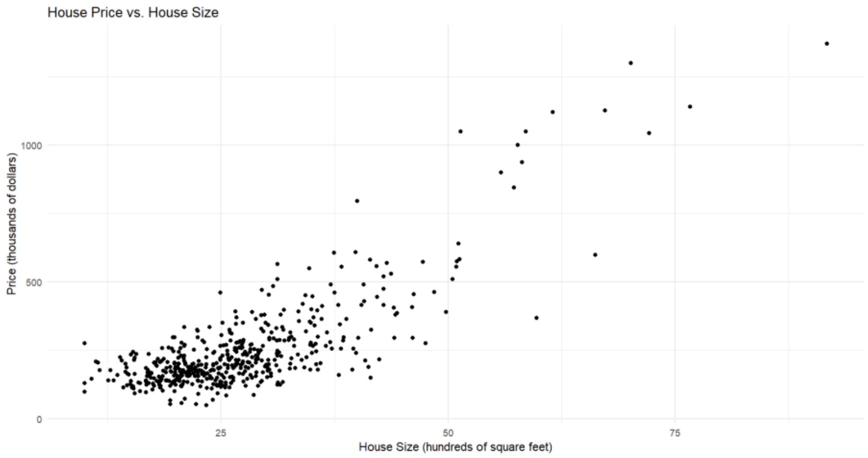
HW0303

Question 17

1. Code: Please check the corresponding R file.

2. Results:

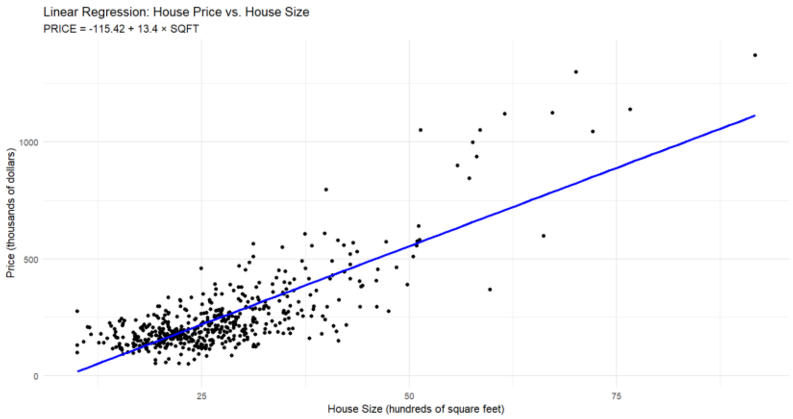
a. Scatter plot of house price against house size created.



b. Linear Model: PRICE = -115.4236 + 13.40294 × SQFT

(se: 13.0882 and 0.4492, respectively)

We estimate that an additional 100 square feet of living area will increase the expected home price by $13,402.94 holding all else constant. The estimated intercept −115.4236 would imply that a house with zero square feet has an expected price of $−115,423.60.

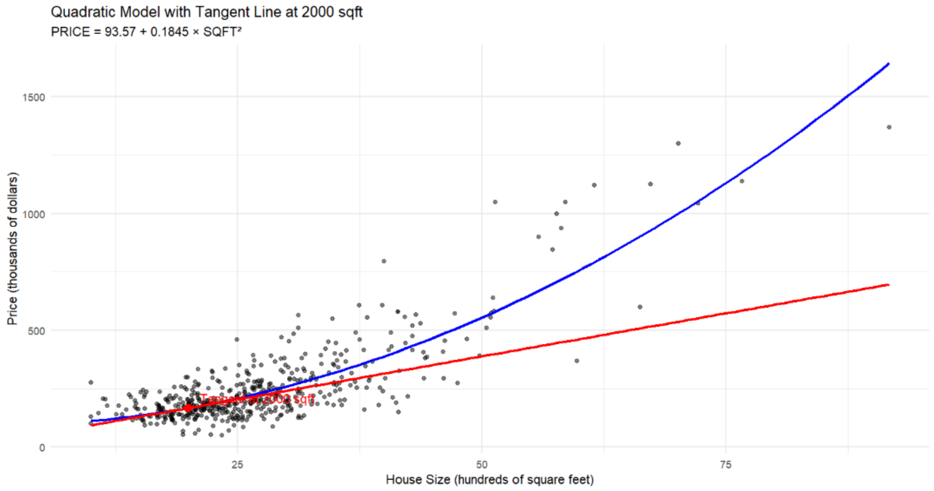


c. Quadratic Model: PRICE = 93.57 + 0.1845 × $SQFT^2$

(se: 6.0722 and 0.00525, respectively)

We estimate that an additional 100 square feet of living area for a 2000 square foot home will increase the expected home price by $7,380.80 holding all else constant.

d. Quadratic model with tangent line at 2000 square feet plotted.

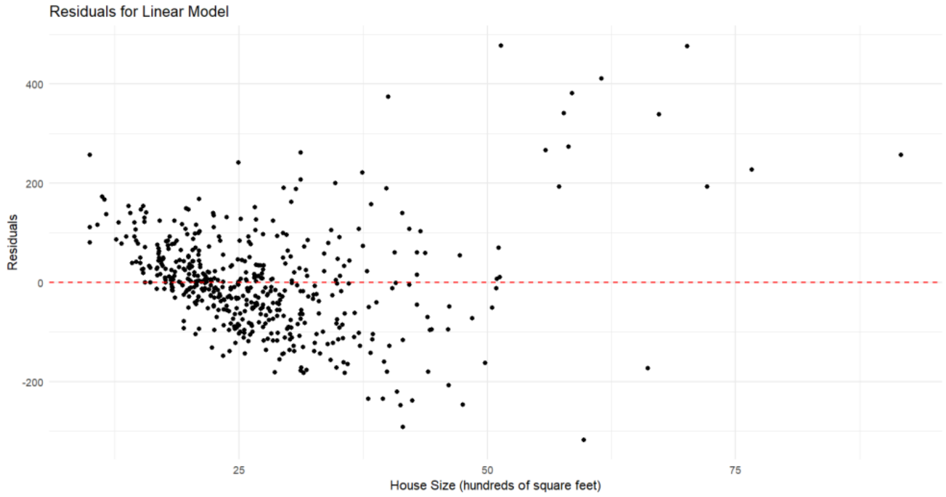


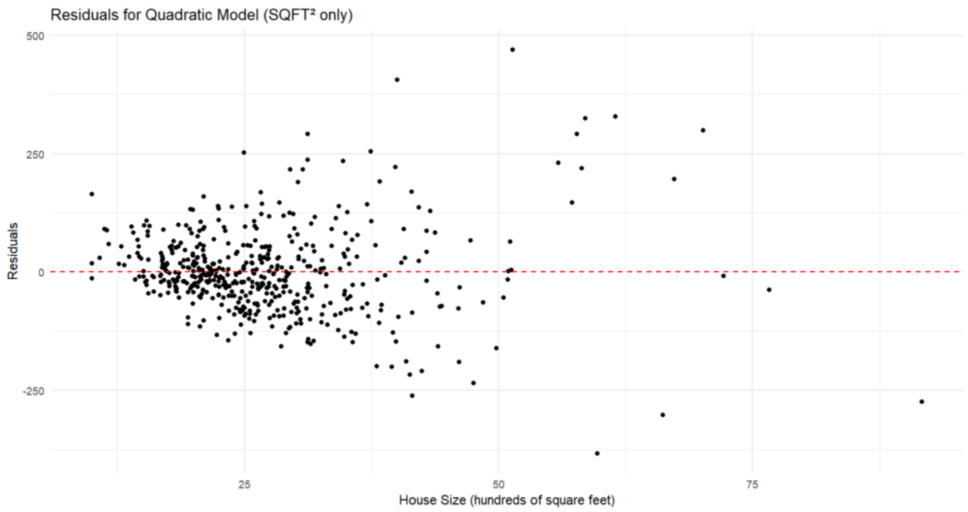
e. Elasticity at 2000 square feet: 0.882

Interpretation: A 1% increase in house size at 2000 square feet results in approximately a 0.882% increase in house price.

f. Residual plots created for both models.

Examining the residual plots for patterns: In both models, the residual patterns do not appear random. The variation in the residuals increases as SQFT increases, suggesting that the homoskedasticity assumption may be violated.



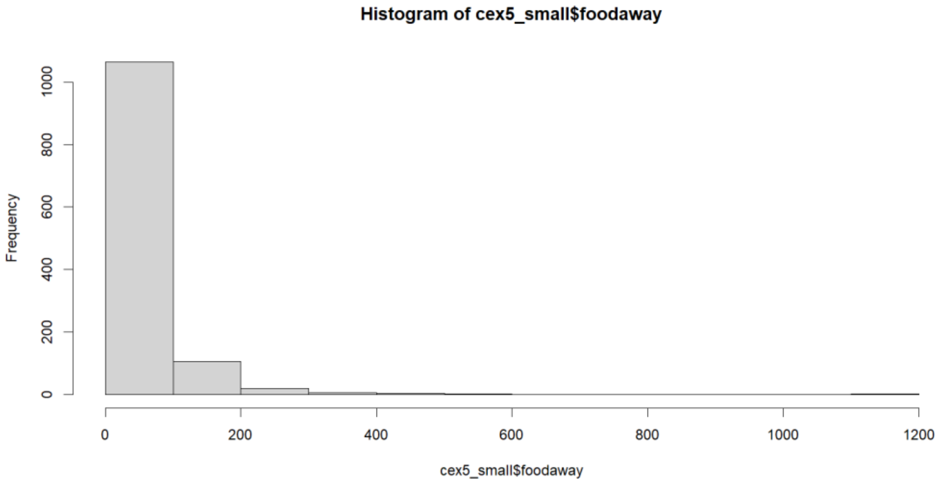


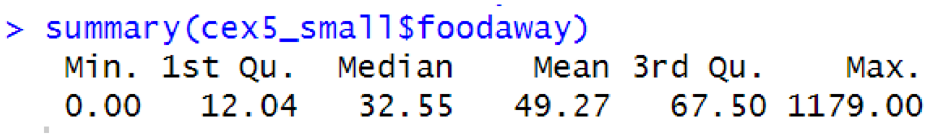
g. Sum of Squared Errors (SSE):

The sum of square residuals linear relationship is 5,262,846.9. The sum of square residuals for the quadratic relationship is 4,222,356.3. In this case the quadratic model has the lower SSE. The lower SSE means that the data values are closer to the fitted line for the quadratic model than for the linear model.

Question 25

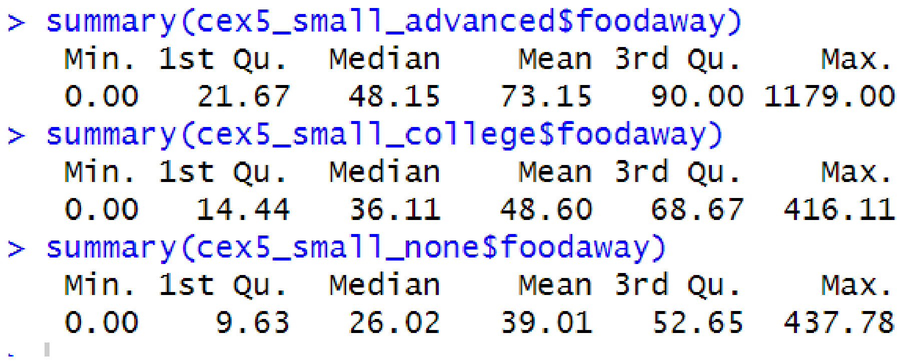
a.





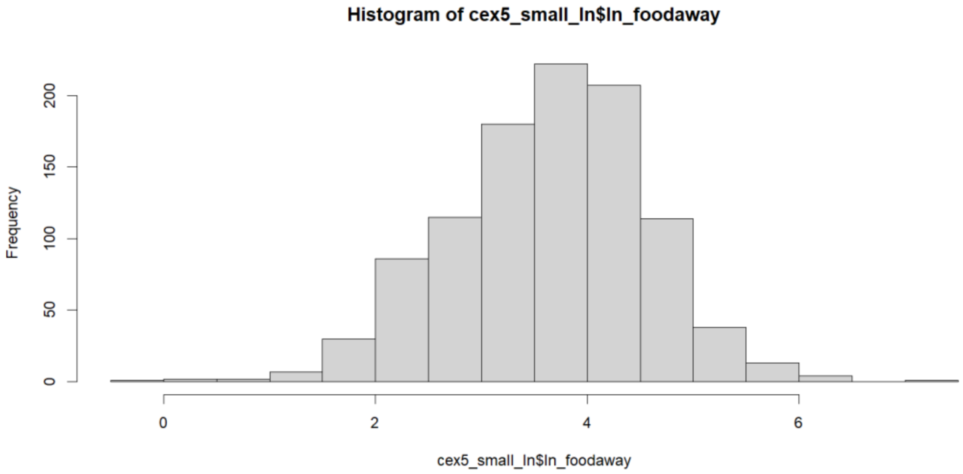
```
> summary(cex5_small$foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   12.04   32.55   49.27   67.50 1179.00
```

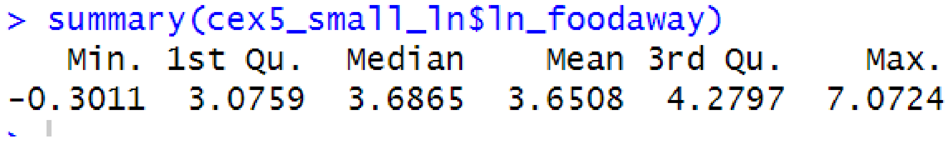
b.



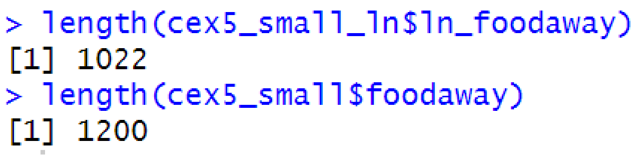
```
> summary(cex5_small_advanced$foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   21.67   48.15   73.15   90.00 1179.00 
> summary(cex5_small_college$foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   14.44   36.11   48.60   68.67  416.11 
> summary(cex5_small_none$foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00    9.63   26.02   39.01   52.65  437.78 
```

c.





```
> summary(cex5_small_ln$ln_foodaway)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-0.3011  3.0759  3.6865  3.6508  4.2797  7.0724 
```



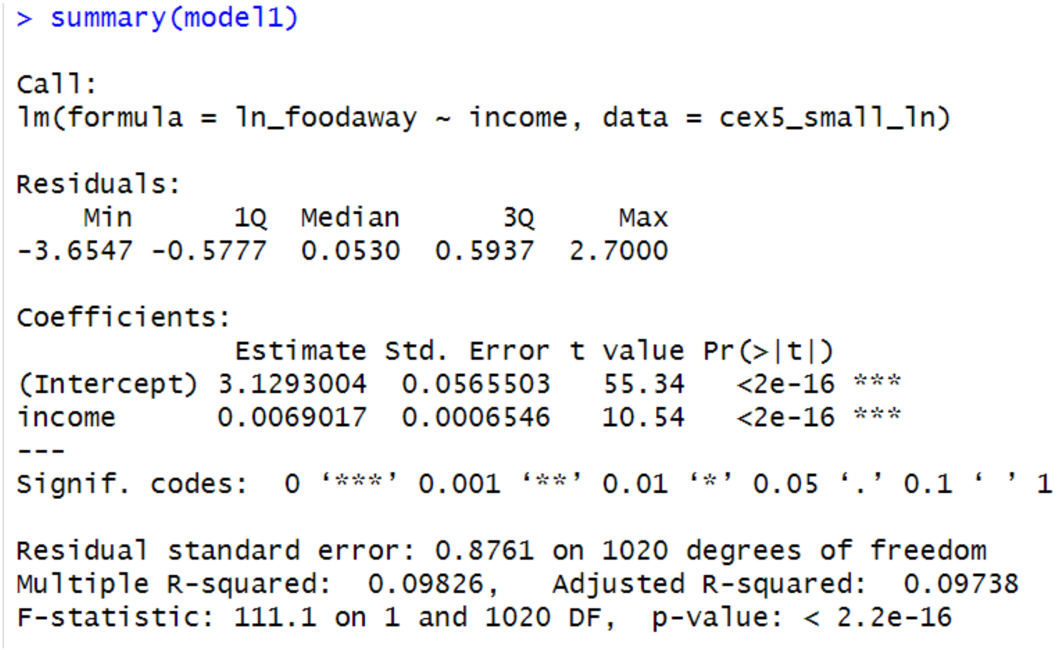
```
> length(cex5_small_ln$ln_foodaway)
[1] 1022
> length(cex5_small$foodaway)
[1] 1200
```

FOODAWAY and ln(FOODAWAY) have different numbers of observations because taking the natural logarithm requires FOODAWAY to be greater than zero. In the dataset, there are 178 observations (1200 - 1022) where FOODAWAY is zero, making it impossible to compute ln(FOODAWAY) for these cases.

d. The estimated model is

$$\ln(\widehat{FOODAWAY}) = 3.1293 + 0.0069 INCOME$$

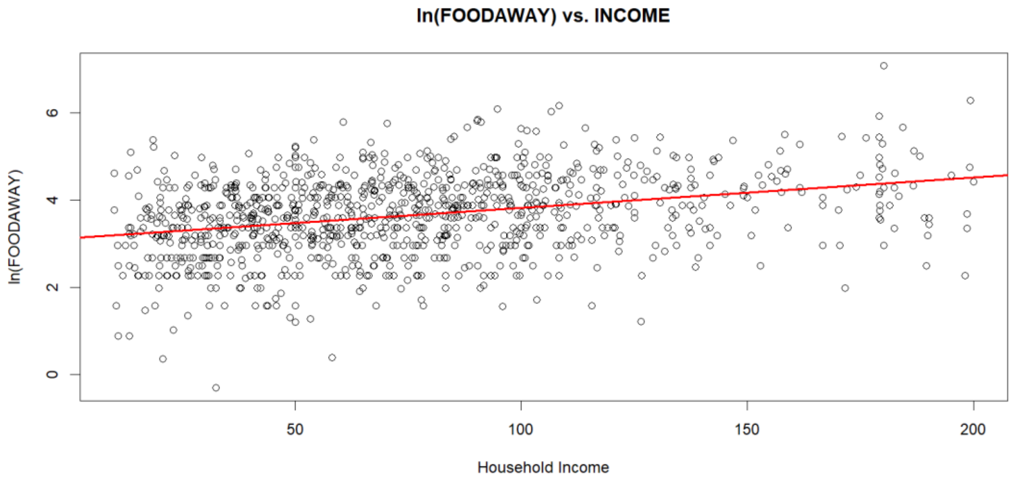
$$(se) \qquad (0.0566) \quad (0.0007)$$



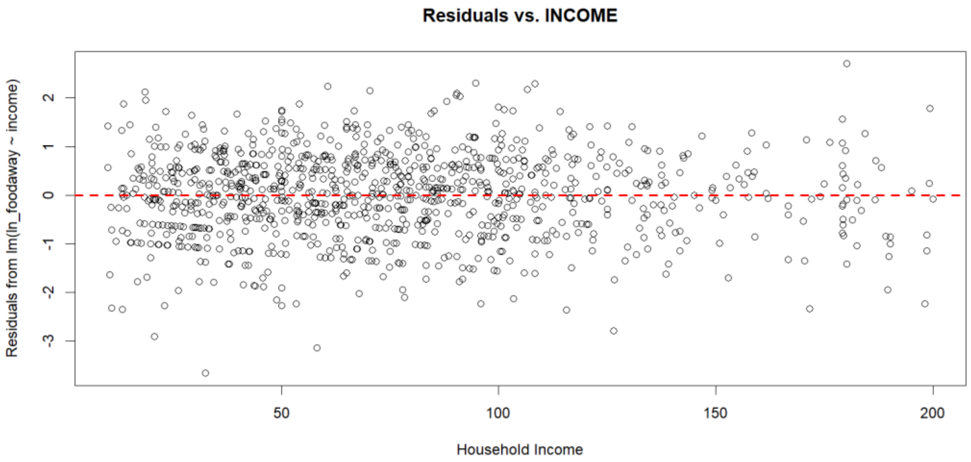
```
> summary(model1)

Call:
lm(formula = ln_foodaway ~ income, data = cex5_small_ln)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6547 -0.5777  0.0530  0.5937  2.7000 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.1293004  0.0565503   55.34   <2e-16 ***
income      0.0069017  0.0006546   10.54   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8761 on 1020 degrees of freedom
Multiple R-squared:  0.09826,	Adjusted R-squared:  0.09738 
F-statistic: 111.1 on 1 and 1020 DF,  p-value: < 2.2e-16
```

Interpreting the estimated slope: We estimate that each additional $100 household income increases food away expenditures per person of about 0.69%, other factors held constant.

e. The plot shows a slight positive association between ln(FOODAWAY) and INCOME.



f. The OLS residuals do appear randomly distributed with no obvious patterns. There are fewer observations at higher incomes, so there is more “white space.”

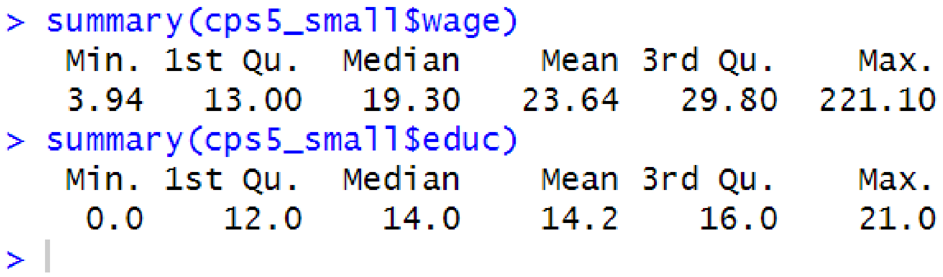


Question 28

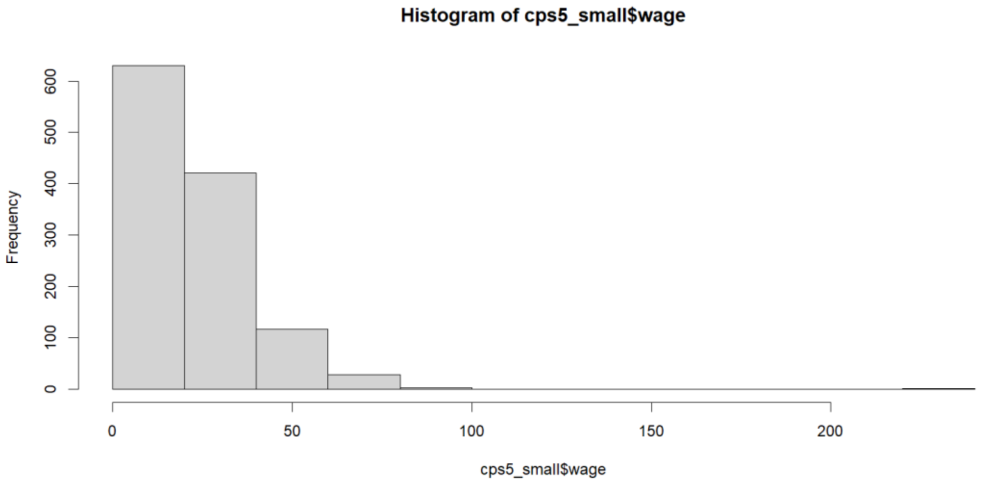
a. The wage data exhibits a pronounced right skew, with most values clustering below $30 and a few extreme outliers pushing the maximum up to $221.10, as reflected by a median wage of around $19.30 and a mean of approximately $23.64.

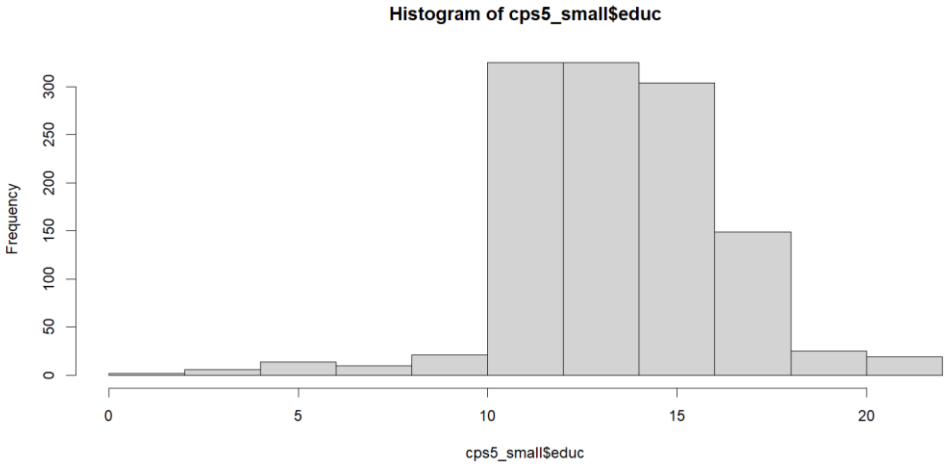
In contrast, education levels are largely concentrated between 12 and 16 years, indicating a unimodal distribution centered on typical schooling durations, with very few observations at the lower and upper extremes.

These characteristics suggest that while the wages might benefit from a log transformation to account for their skewness in modeling, the education variable appears well-suited for standard analytical approaches.



```
> 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 
> 
```

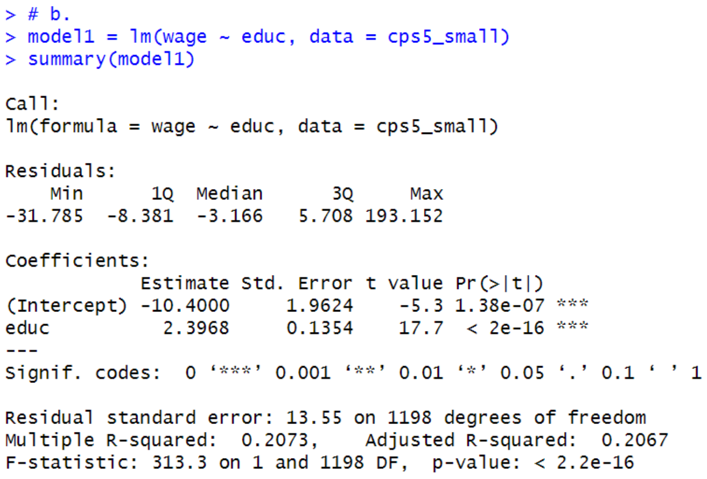




b. The estimated model is:

$$\widehat{\text{WAGE}} = -10.4000 + 2.3968\,\text{EDUC}$$

$$(\text{se}) \quad (1.9624) \qquad (0.1354)$$



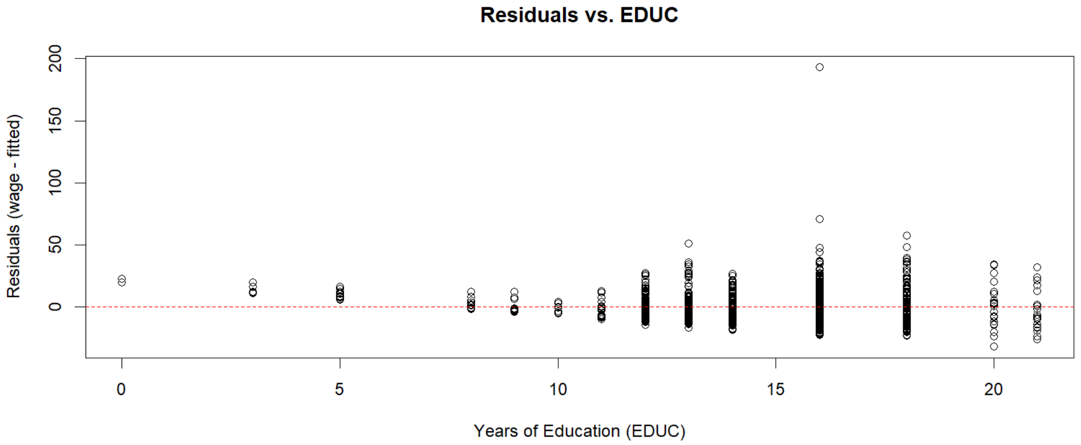
```
> # b.
> model1 = lm(wage ~ educ, data = cps5_small)
> summary(model1)

Call:
lm(formula = wage ~ educ, data = cps5_small)

Residuals:
    Min      1Q  Median      3Q     Max 
-31.785  -8.381  -3.166   5.708 193.152 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -10.4000     1.9624    -5.3 1.38e-07 ***
educ          2.3968     0.1354    17.7  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.55 on 1198 degrees of freedom
Multiple R-squared:  0.2073,	Adjusted R-squared:  0.2067 
F-statistic: 313.3 on 1 and 1198 DF,  p-value: < 2.2e-16
```

The model suggests that, for each additional year of education, the hourly wage is expected to rise by $2.3968, holding everything else constant. Regarding the intercept (-10.4000), this literally means that if someone had zero years of education (i.e., EDUC = 0), the model would predict a wage of -$10.40.

c. The residuals-versus-education plot does not exhibit a clear systematic pattern, such as a "U" shape, indicating that a linear specification in EDUC alone is not grossly violated.

The residuals are generally centered around zero, though a few high positive residuals (notably around EDUC = 15 and 20) suggest some individuals earn more than predicted. There is a slight increase in the spread of residuals at higher education levels, hinting at mild heteroskedasticity but not to an extreme degree.

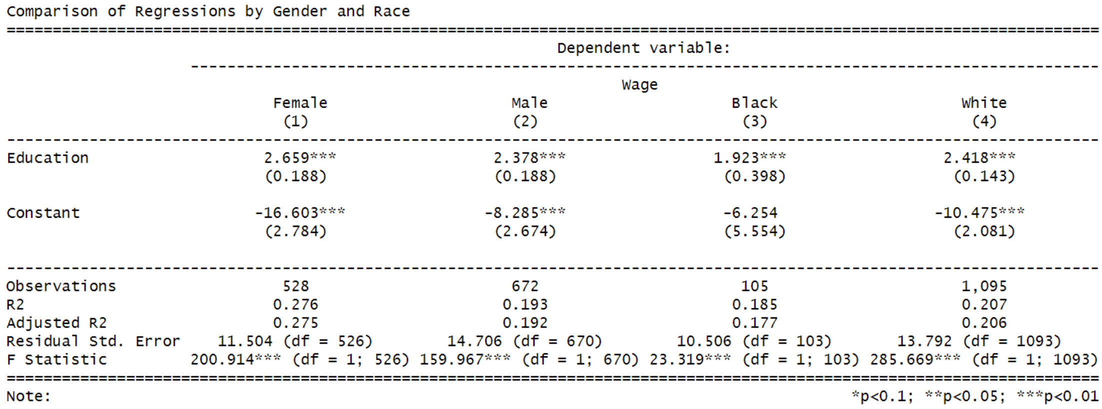
The assumptions of strict exogeneity, linearity, and uncorrelated errors appear reasonably satisfied, and there is sufficient variation in EDUC. Overall, the residuals display a relatively random scatter around zero, which aligns with classical OLS assumptions (SR1–SR5) and suggests that the linear model is appropriate with no severe violations.



d. The regressions of **wage** on **education** across demographic subgroups reveal notable differences in returns to schooling. **Females** exhibit the highest return per year of education ($2.66), followed by **males** ($2.38) and **whites** ($2.42), while **blacks** have the lowest return ($1.92).

The negative intercepts, though unrealistic for zero years of education, simply reflect the extrapolated best-fit line. In terms of explanatory power, **females** have the highest R-squared (0.276), indicating that education alone accounts for more wage variation in this group, whereas **blacks** have both the lowest return to education and the lowest R-squared (0.185), suggesting other factors significantly influence their wages.

Overall, these findings highlight that the linear relationship between education and wages varies across groups, with females benefiting the most from additional schooling, blacks the least, and males and whites falling in between.



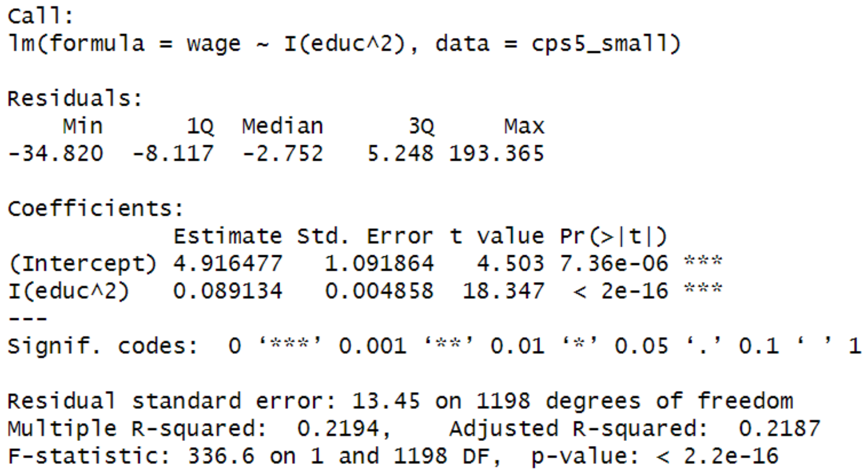
```
Comparison of Regressions by Gender and Race
=======================================================================================================================
                                                            Dependent variable:                                        
                    ---------------------------------------------------------------------------------------------------
                                                                   Wage                                                
                             Female                    Male                    Black                    White          
                              (1)                      (2)                      (3)                      (4)           
-----------------------------------------------------------------------------------------------------------------------
Education                   2.659***                 2.378***                1.923***                 2.418***         
                            (0.188)                  (0.188)                  (0.398)                  (0.143)         
                                                                                                                       
Constant                   -16.603***               -8.285***                 -6.254                 -10.475***        
                            (2.784)                  (2.674)                  (5.554)                  (2.081)         
                                                                                                                       
-----------------------------------------------------------------------------------------------------------------------
Observations                  528                      672                      105                     1,095          
R2                           0.276                    0.193                    0.185                    0.207          
Adjusted R2                  0.275                    0.192                    0.177                    0.206          
Residual Std. Error    11.504 (df = 526)        14.706 (df = 670)        10.506 (df = 103)       13.792 (df = 1093)    
F Statistic         200.914*** (df = 1; 526) 159.967*** (df = 1; 670) 23.319*** (df = 1; 103) 285.669*** (df = 1; 1093)
=======================================================================================================================
Note:                                                                                       *p<0.1; **p<0.05; ***p<0.01
```

e. The estimated model is:

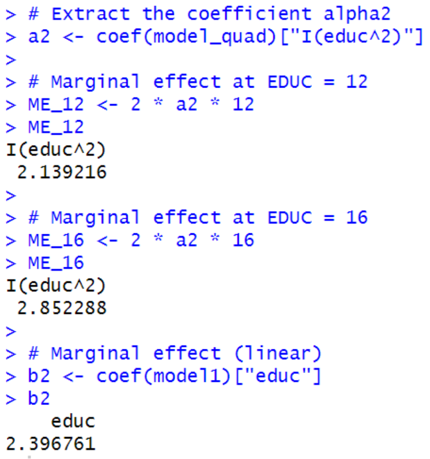
$$\widehat{\text{WAGE}} = 4.9165 + 0.08913\,(\text{EDUC})^2.$$

- **Marginal Effect** of another year of education:

$$\frac{\partial\,\widehat{\text{WAGE}}}{\partial\,\text{EDUC}} = 2 \times 0.08913 \times \text{EDUC} = 0.17826 \times \text{EDUC}.$$



```
Call:
lm(formula = wage ~ I(educ^2), data = cps5_small)

Residuals:
    Min      1Q  Median      3Q     Max 
-34.820  -8.117  -2.752   5.248 193.365 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.916477   1.091864   4.503 7.36e-06 ***
I(educ^2)   0.089134   0.004858  18.347  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.45 on 1198 degrees of freedom
Multiple R-squared:  0.2194,	Adjusted R-squared:  0.2187 
F-statistic: 336.6 on 1 and 1198 DF,  p-value: < 2.2e-16
```



```
> # Extract the coefficient alpha2
> a2 <- coef(model_quad)["I(educ^2)"]
>
> # Marginal effect at EDUC = 12
> ME_12 <- 2 * a2 * 12
> ME_12
I(educ^2)
 2.139216
>
> # Marginal effect at EDUC = 16
> ME_16 <- 2 * a2 * 16
> ME_16
I(educ^2)
 2.852288
>
> # Marginal effect (linear)
> b2 <- coef(model1)["educ"]
> b2
    educ
2.396761
```

So, in the quadratic model, the return to education increases with additional schooling—someone moving from 12 to 13 years of education gains about \$2.14/hour, while someone moving from 16 to 17 years gains around \$2.85/hour.

For the linear model, its marginal effect is a constant 2.40 dollars/hour per extra year of education, regardless of whether you have any years of schooling.

Key Insights:

At 12 years: The quadratic model's marginal effect (about \$2.14) is slightly lower than the linear slope (\$2.40).

At 16 years: The quadratic model's marginal effect (about \$2.85) is higher than the linear slope.

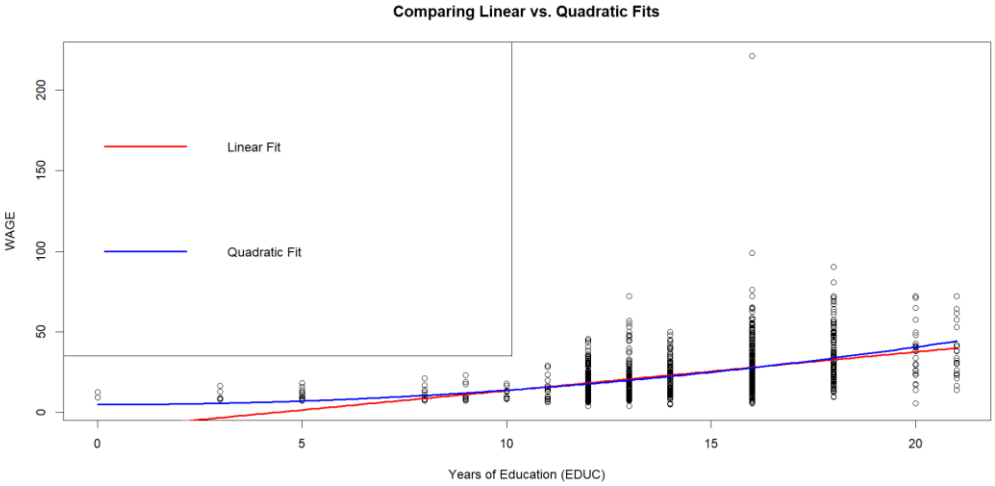
Thus, the quadratic specification implies returns to education grow as schooling increases, whereas the linear model forces a constant return at all education levels.

f. Linear: Adjusted R-squared = 0.2067

Quadratic: Adjusted R-squared = 0.2187

The quadratic model provides a slightly better fit than the linear model, as seen in both the plot and regression results:

1. Visually, it avoids the unrealistic negative wage predictions at low education levels and better captures the steeper wage growth for higher education (15–20 years).
2. Statistically, the quadratic model shows a marginally higher R2 and lower SSE, indicating it explains slightly more wage variation.
3. While the difference is not dramatic, the quadratic specification more flexibly accommodates increasing wage growth with education, rather than imposing a constant slope.


Comparing Linear vs. Quadratic Fits
Linear Fit
Quadratic Fit
WAGE
0
50
100
150
200
0
5
10
15
20
Years of Education (EDUC)