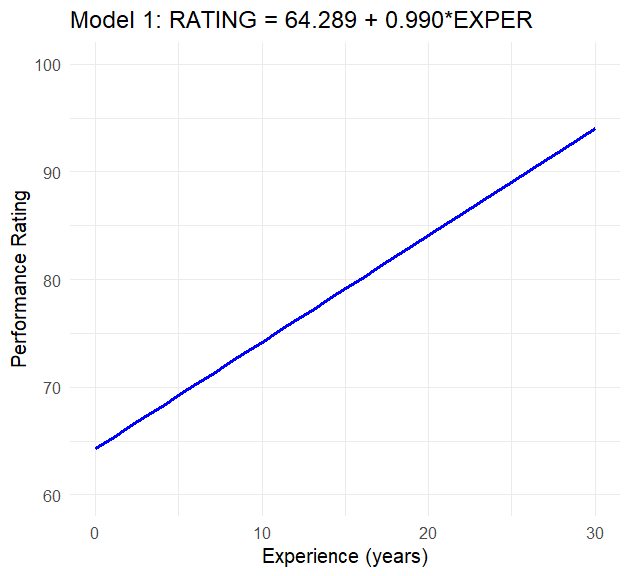
Le Thi Phuong Thao

Student: 413707007

HW0317

**Question 4.4**

1. Plot

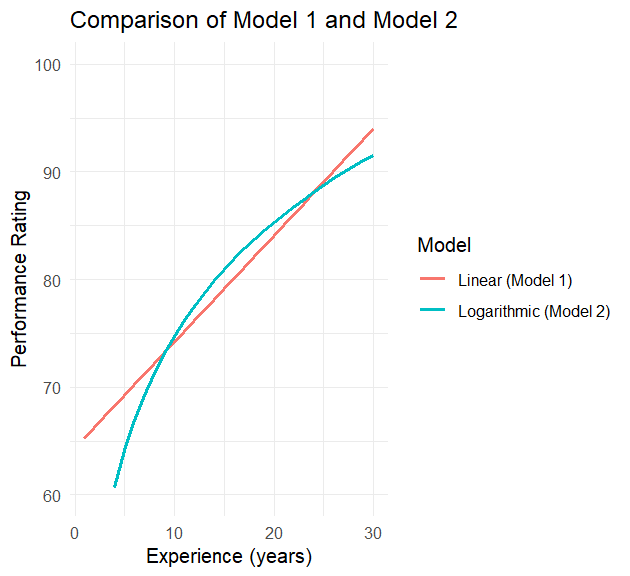
****

1. Plot

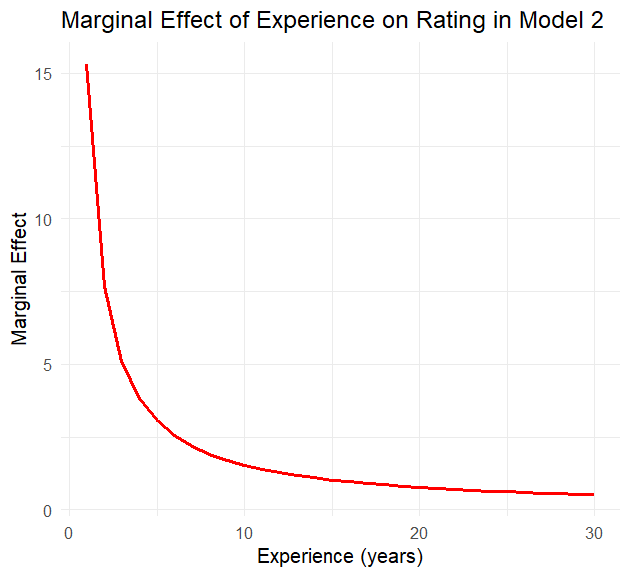
**A graph with a red line

AI-generated content may be incorrect.**

1. Plot

****

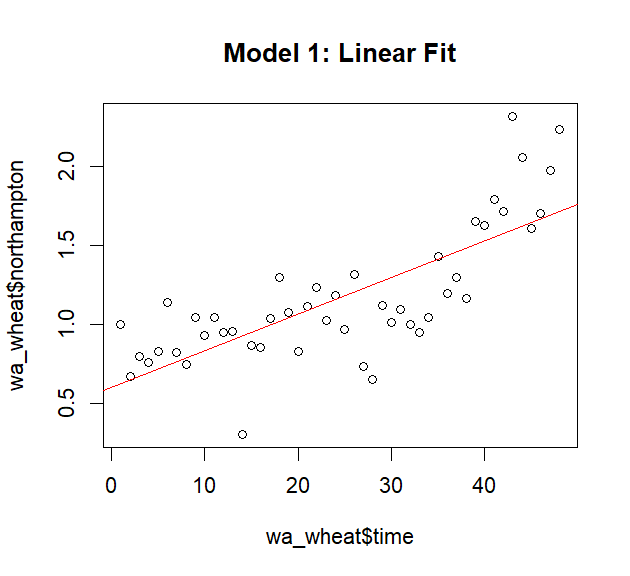
1. Plot



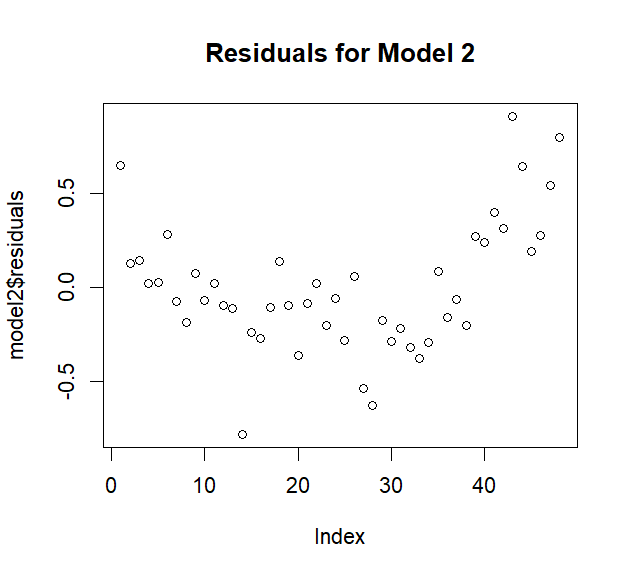
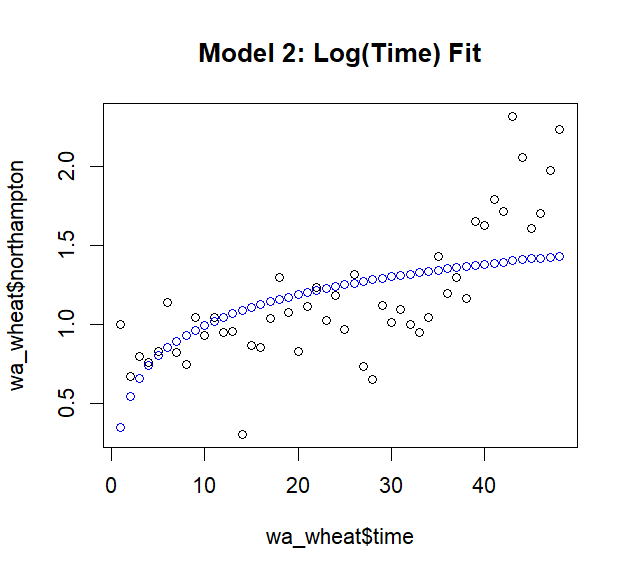
1. Model 2 fits the data better with an R² of 0.6414 compared to Model 1's R² values
2. Model 2 (logarithmic model) is more plausible because it captures diminishing or increasing marginal effects: the effect of a change in x becomes smaller or larger as x increases. Diminishing returns: the first 100 hours may boost skill (and income) more than the 1000th hour

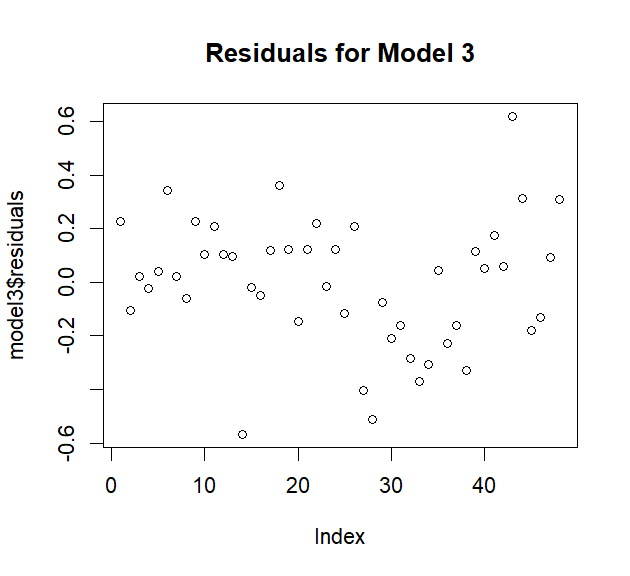
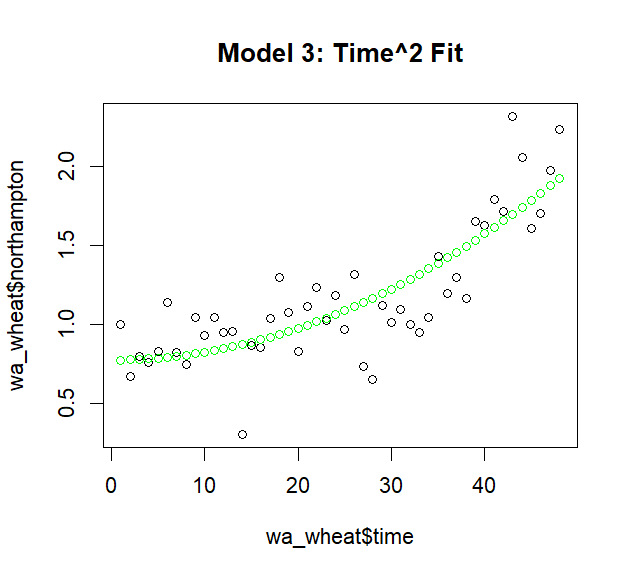
**Question 4.28**

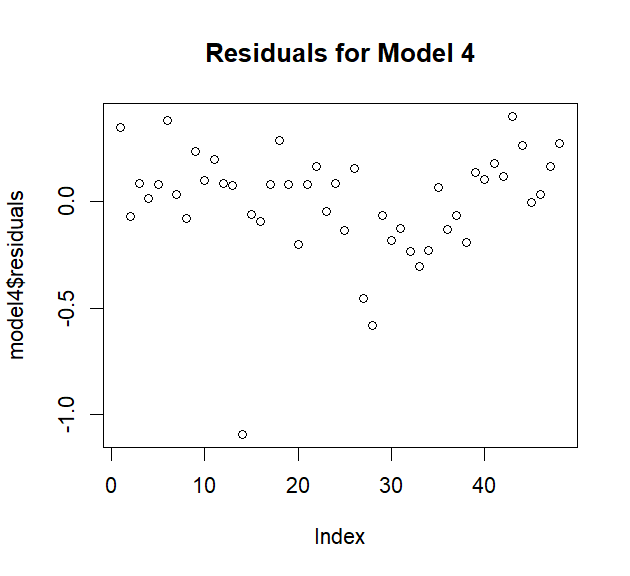
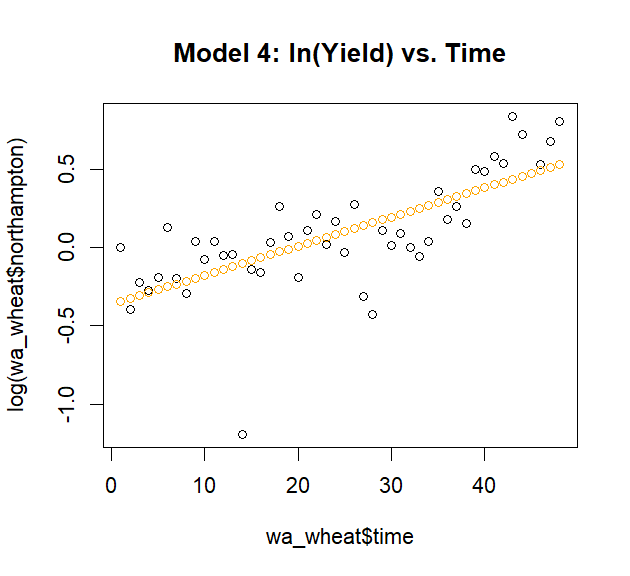
1. Plot

A graph with dots and numbers

AI-generated content may be incorrect.







> print(sw\_test1)

Shapiro-Wilk normality test

data: model1$residuals

W = 0.98236, p-value = 0.6792

> print(sw\_test2)

Shapiro-Wilk normality test

data: model2$residuals

W = 0.96657, p-value = 0.1856

> print(sw\_test3)

Shapiro-Wilk normality test

data: model3$residuals

W = 0.98589, p-value = 0.8266

> print(sw\_test4)

Shapiro-Wilk normality test

data: model4$residuals

W = 0.86894, p-value = 7.205e-05

> print(paste("R-squared for Model 1:", R2\_1))

[1] "R-squared for Model 1: 0.577836870356481"

> print(paste("R-squared for Model 2:", R2\_2))

[1] "R-squared for Model 2: 0.338573313180548"

> print(paste("R-squared for Model 3:", R2\_3))

[1] "R-squared for Model 3: 0.6890100869004"

> print(paste("R-squared for Model 4:", R2\_4))

[1] "R-squared for Model 4: 0.50735657110073"

Model 3 seems to be preferred, it has the largest R squared, and normally error

1. Intercept model 3: for every one-unit increase in time squared term, the Northampton variable increases by about 0.0004986 units, assuming other factors remain constant.
2. Using your chosen specification, identify any unusual observations, based on the studentized residuals, LEVERAGE, DFBETAS, and DFFITS

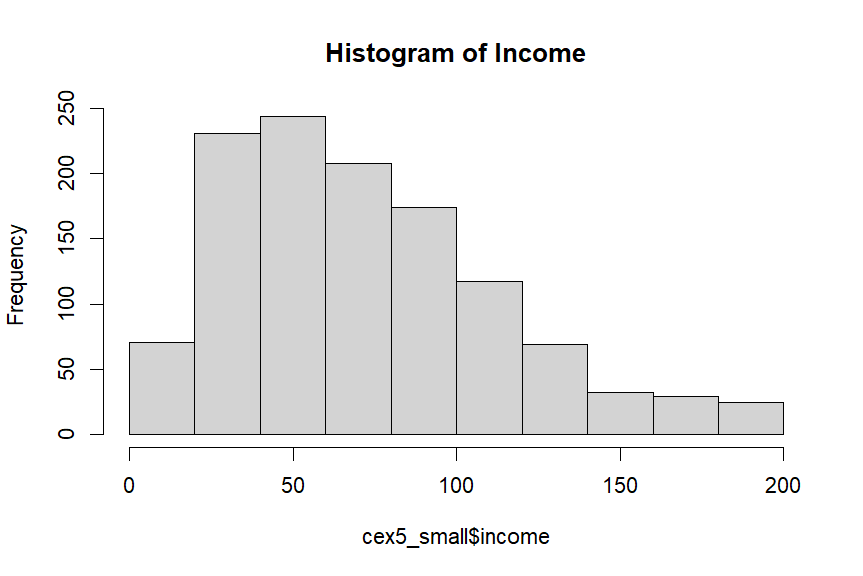
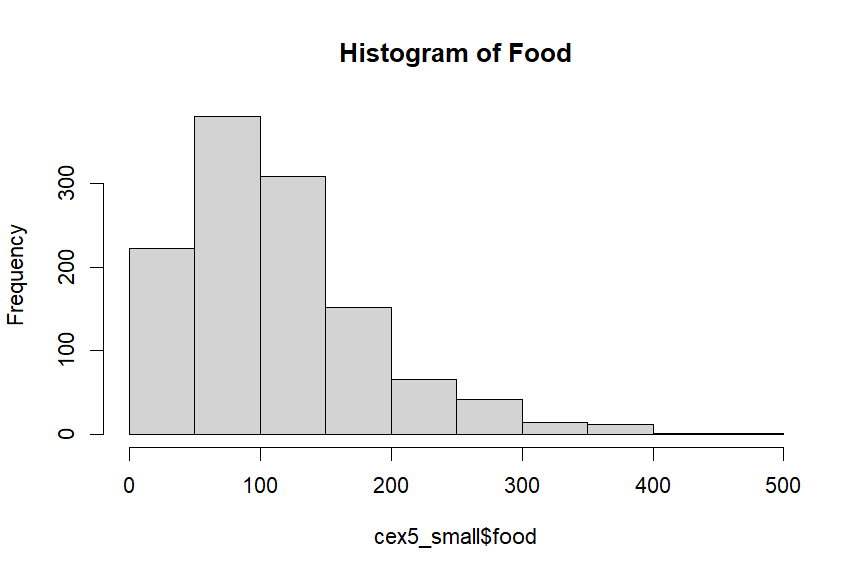
A computer code with text

AI-generated content may be incorrect.

**Question 4.29**

1. Answer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Var. | mean | median | min | max | Std |
| Food | 114.44 | 99.80 | 9.63 | 114.44 | 72.6575 |
| Income | 72.14 | 65.29 | 10.00 | 200.00 | 41.65228 |



A screenshot of a computer code

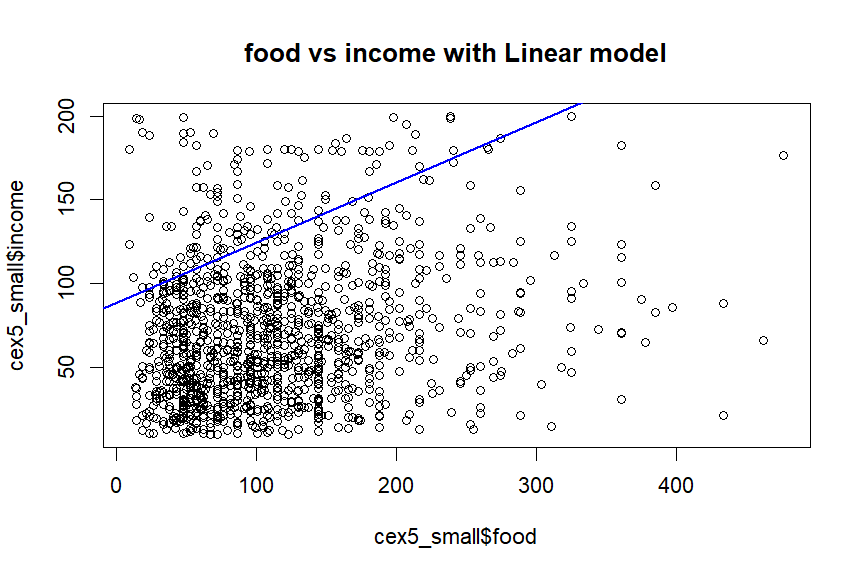
AI-generated content may be incorrect.

Both distributions are positively skewed with mean’s greater than medians. They are not symmetrical or bell shaped curves. For INCOME the Jarque-Bera statistic is 148.21 and for FOOD expenditure it is 648.65 with a significant p-value < 1%. We reject the null hypothesis of normality for each variable.

1. Estimate the linear relationship FOOD = β1 + β2INCOME + e.

A screenshot of a computer code

AI-generated content may be incorrect.



A number and numbers on a white background

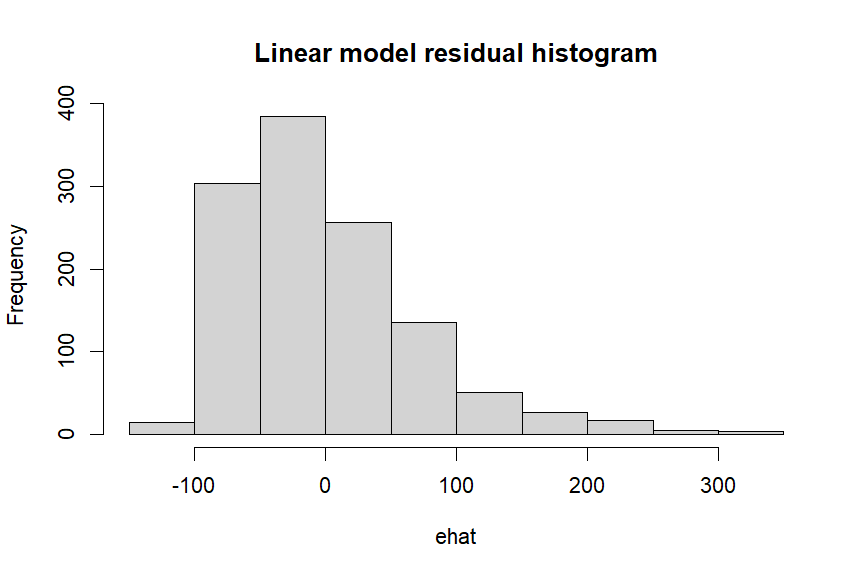
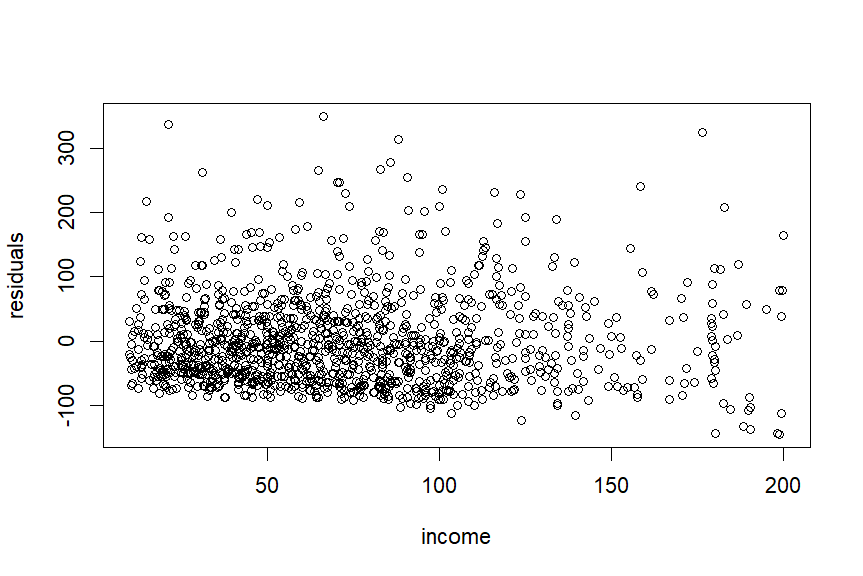
AI-generated content may be incorrect.

The 95% interval estimate of the slope is [0.2619, 0.4555].

We have not estimated the effect of changing income on average FOOD relatively precisely, the fitted line does not capture almost the density of the plot

1. The least squares residuals are plotted as below

The positive skew at each income is clear. There is not a clear “spray” pattern except at high incomes. The residual histogram shows the skewness. The Jarque-Bera statistic is 624.19, P-value is significant at 1% level so residual is not normally distributed.



A black text on a white background

AI-generated content may be incorrect.

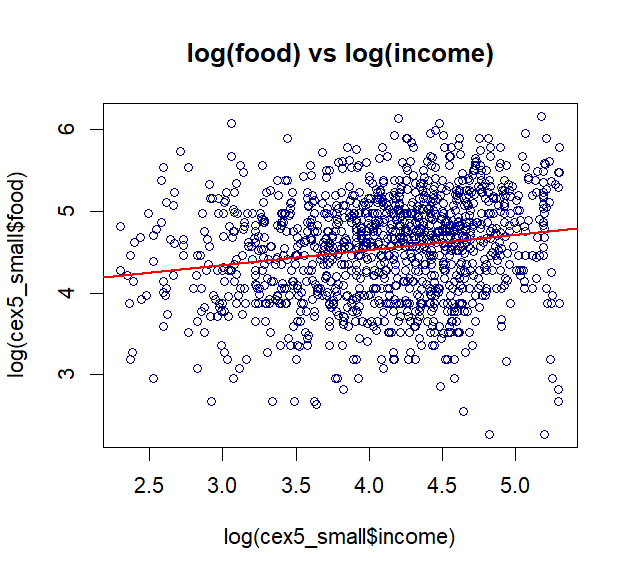
1. Calculate both a point estimate and a 95% interval estimate of the elasticity of food expenditure with respect to income at INCOME = 19, 65, and 160,

A white background with black numbers and letters

AI-generated content may be incorrect.

The estimated elasticities are dissimilar, as the confidence intervals for the three models do not overlap. The elasticity initially increases with income. Ideally, at higher income levels, food expenditure no longer increases significantly. This data seem to cover low-income data.

1. Plot



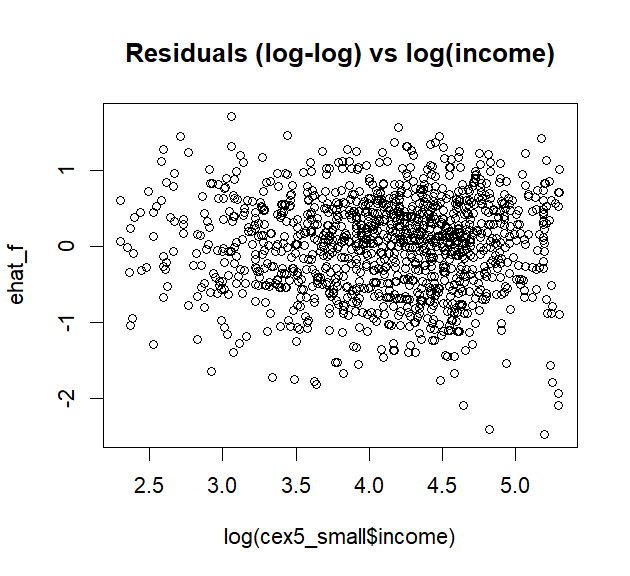
The generalized R2 is 0.03965 which is slightly smaller than the R2 from the linear model, however, the fitted line is better well-defined than the linear-linear model

1. The 95% interval estimate of the elasticity (beta 2) [0.1293432, 0.2432675]

In part d, the 95% interval estimate of the slope is [0.2619, 0.4555]. The elasticity of

food expenditure from the log-log model is similar and within the range of the linear-linear model. With income = 65, log-log model has elasticity = 0.18631 while linear-linear is Beta2 X/y = 0.20838756

1. The residual scatter from the log-log model

A graph of a number of gray bars

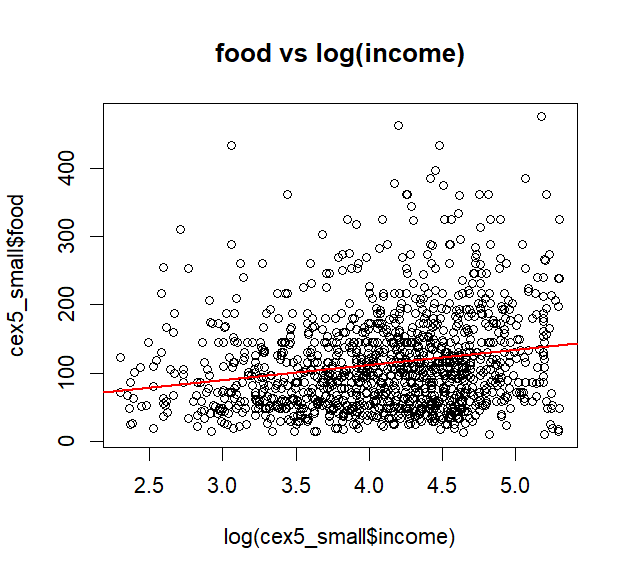
AI-generated content may be incorrect.

A white background with black text

AI-generated content may be incorrect.

The Jarque- Bera statistic is 25.85 which is greater than the 5% critical value 5.99. So, we reject the null hypothesis that the log-log regression errors are normal.

1. the linear-log relationship FOOD = α1 + α2ln(INCOME) + e



The figure is much like that for the linear model, and not as well defined as that for the log-log model. The R2 = 0.038, which is smaller than that of the linear model, and smaller than the generalized R2 from the log-log model

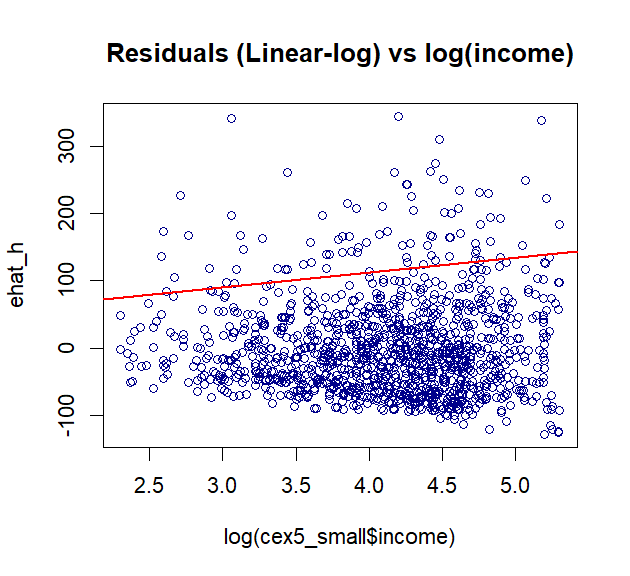
1. Ans.

A white background with black numbers and letters

AI-generated content may be incorrect.

the elasticity of food expenditure is dissimilar to those from the other models

1. Ans: The Jarque-Bera statistic is 628.07 which is far greater than the 5.99 critical value. We reject the normality of the model errors. The data scatter suggests a slight “spray” pattern



A black text on a white background

AI-generated content may be incorrect.

k) The linear model is counter-intuitive with increasing income elasticity. The linear-log model certainly satisfies economic reasoning, but the residual pattern is not an ideal random scatter.

The log-log model implies that the income elasticity is constant for all income levels, which is not impossible to imagine, and the residual scatter is the most random, and the residuals are the least non-normal, based on skewness and kurtosis. On these grounds the log-log model seems like a good choice