**Full name:** Nguyen Nhut Vu Truong

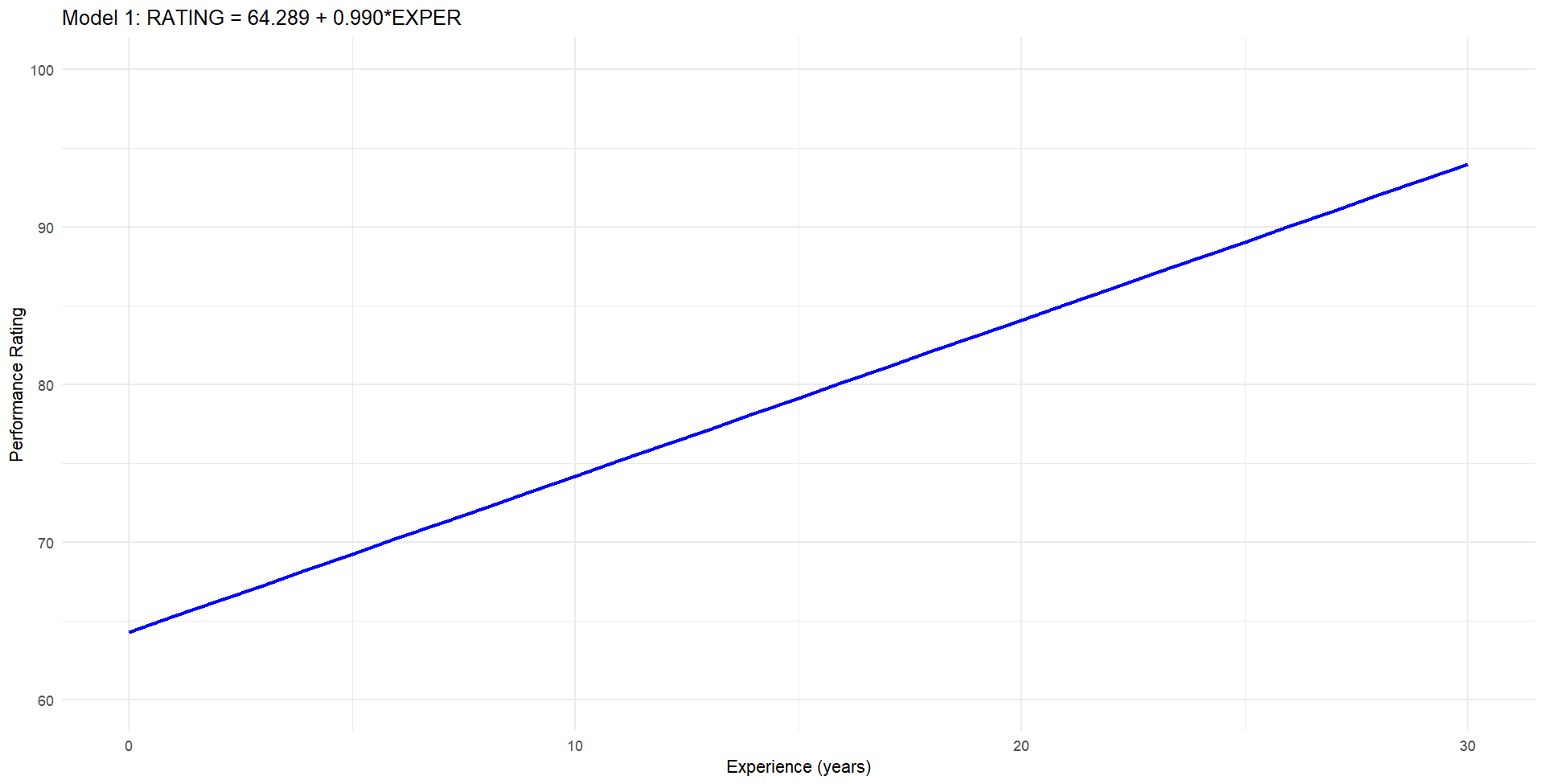
**Student ID:** 413707008

**Course:** Financial Econometrics

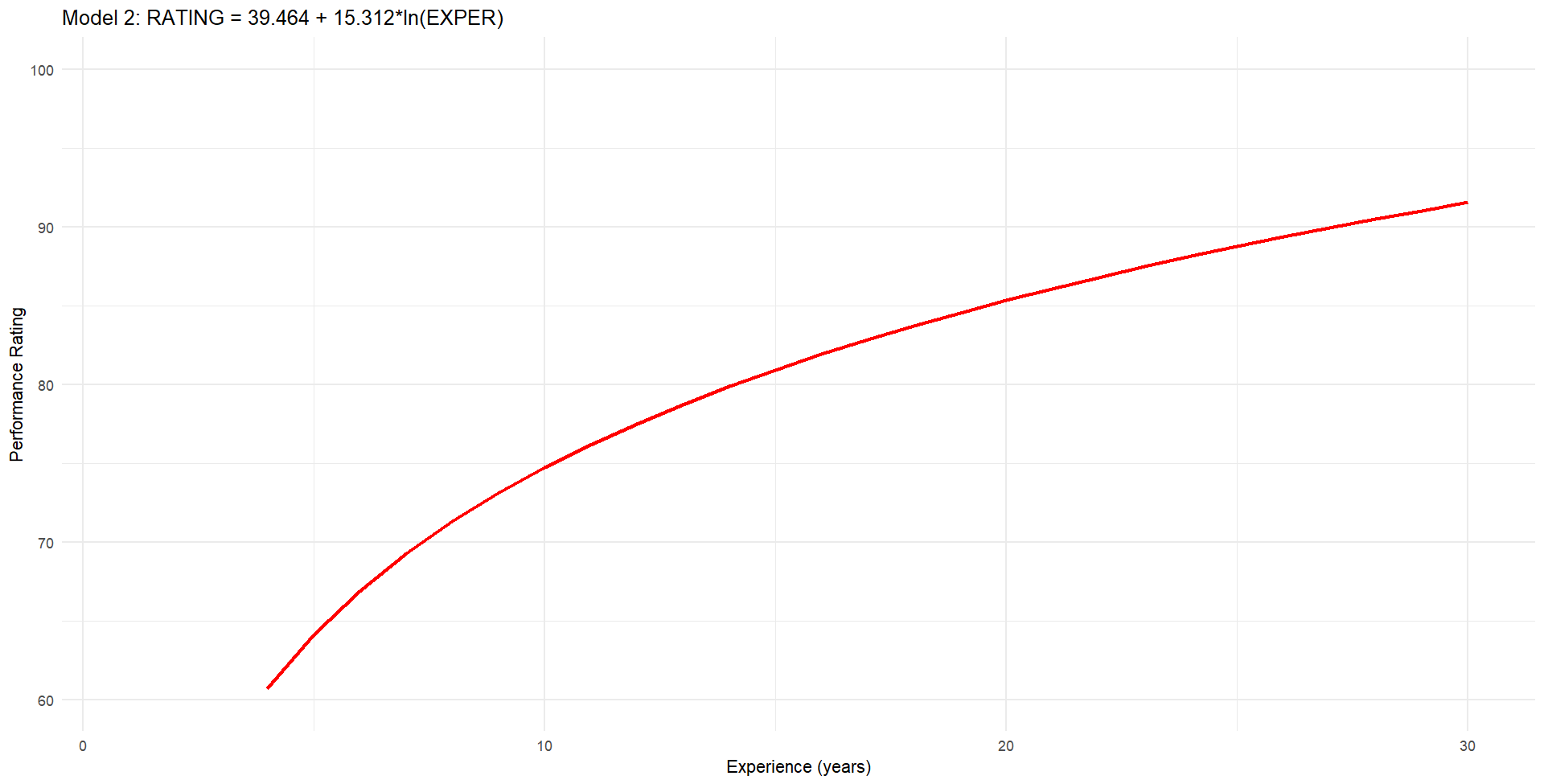
**HW0317**

# **Question 4**

a.



b.



Why four artists with no experience are excluded from Model 2: The natural logarithm of zero (ln(0)) is undefined. Since Model 2 uses ln(EXPER), artists with zero experience cannot be included in the estimation.

c.

In Model 1, the marginal effect is the coefficient of EXPER, which is constant at 0.990 regardless of experience level:

(i) For an artist with 10 years of experience: 0.990

(ii) For an artist with 20 years of experience: 0.990

This means each additional year of experience increases the expected rating by 0.990 points, regardless of how much experience the artist already has.

d.

(i) For an artist with 10 years of experience: 15.312 × (1/10) = 1.5312

(ii) For an artist with 20 years of experience: 15.312 × (1/20) = 0.7656

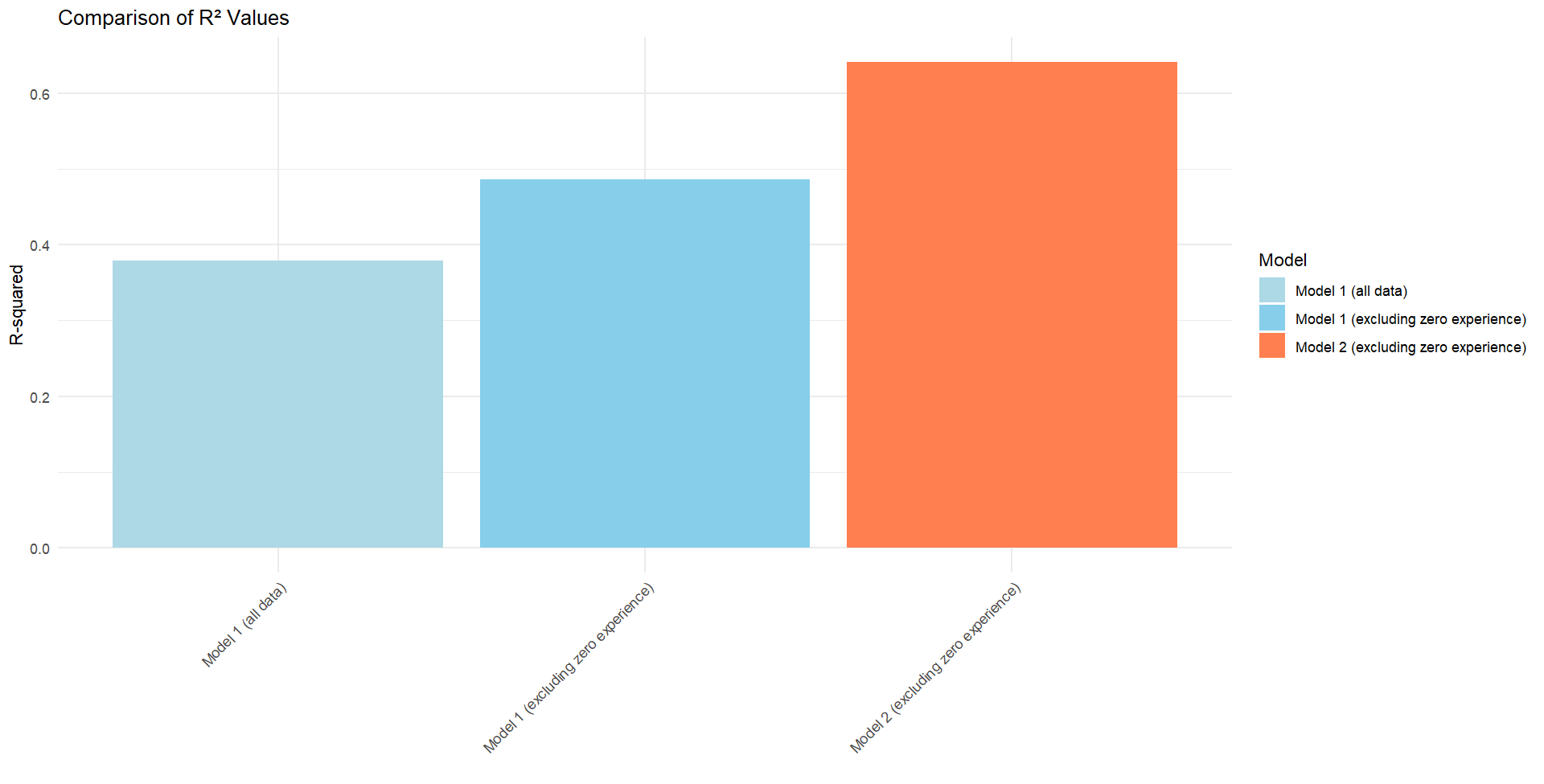
In Model 2, the marginal effect decreases as experience increases, showing that each additional year provides less benefit for more experienced artists.

e.

To evaluate which model fits the data better, we compare their R² values:

* Model 1 (all data): R² = 0.3793
* Model 1 (only artists with experience): R² = 0.4858
* Model 2: R² = 0.6414

Model 2 has the highest R² value (0.6414), indicating it explains about 64% of the variation in performance ratings, compared to only about 38-49% for Model 1. This suggests Model 2 provides a better fit to the observed data.

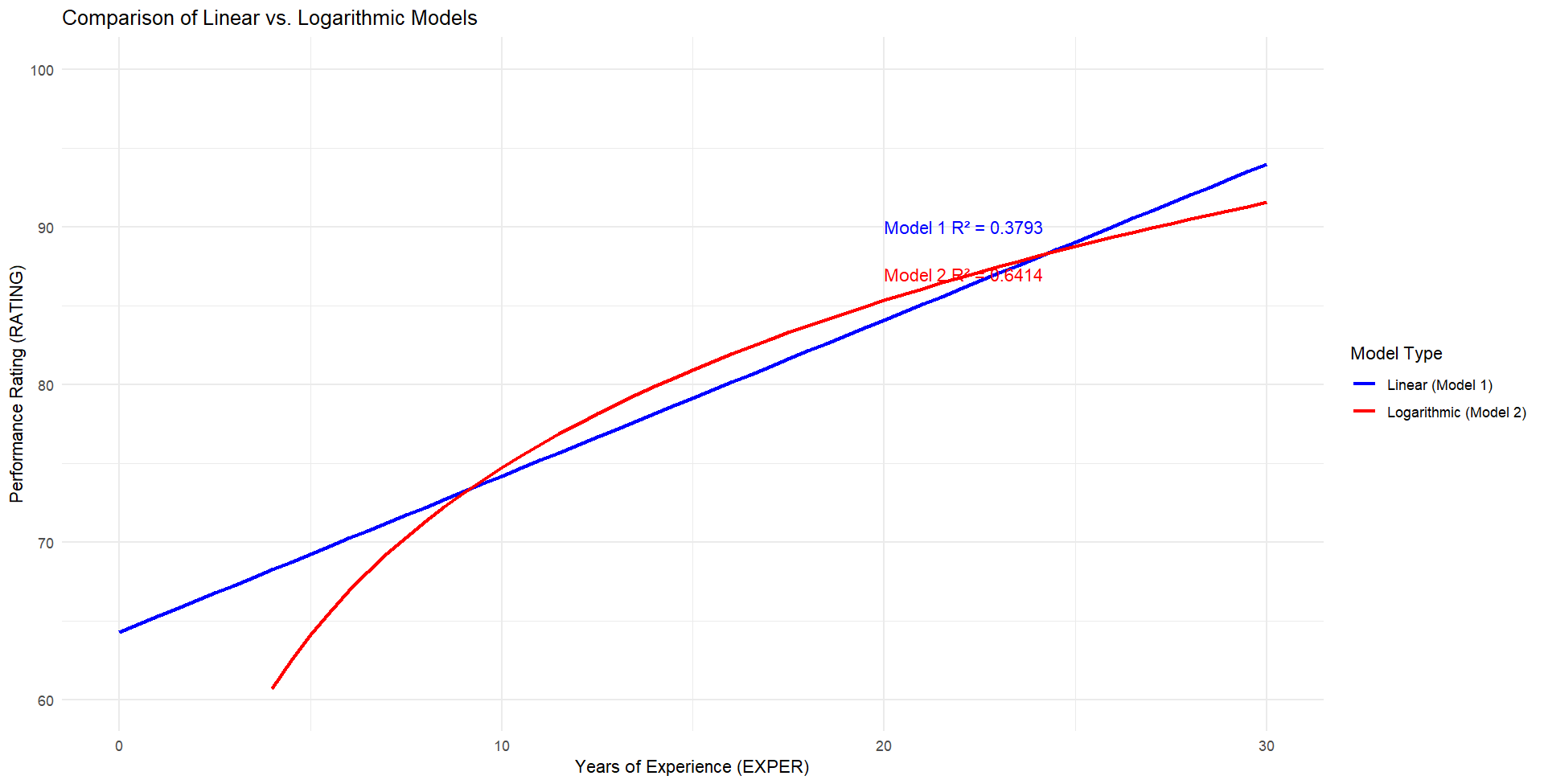


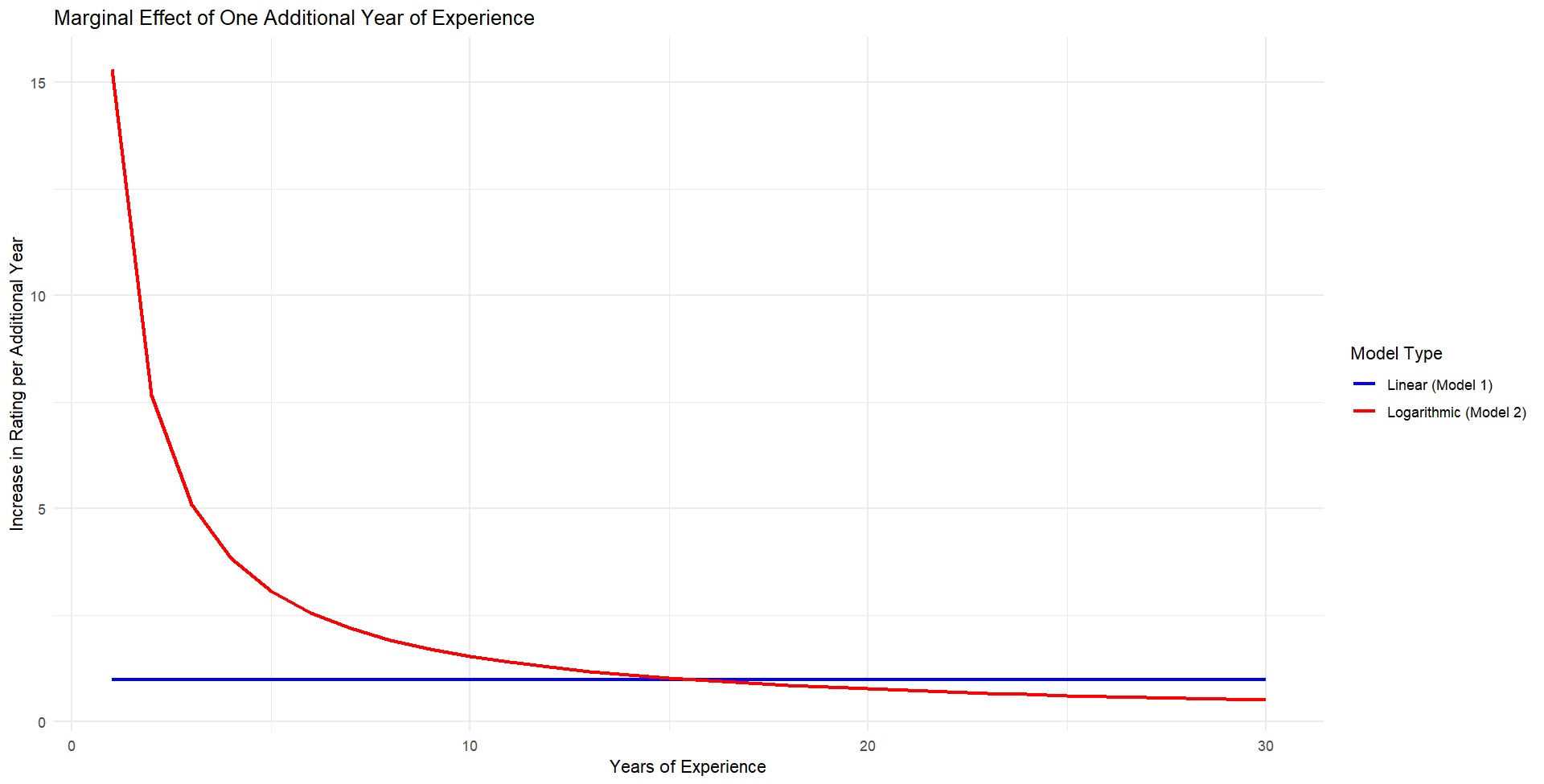
f.

From an economic perspective, Model 2 (logarithmic) is more plausible for several reasons:

1. **Diminishing returns**: In most professions, the effect of additional experience typically diminishes over time. Model 2 captures this economic principle with its logarithmic form, while Model 1 unrealistically implies constant returns to experience.
2. **Learning curve**: Professionals generally improve rapidly early in their careers and then plateau. Model 2 correctly shows steeper improvements for less experienced artists that gradually level off.
3. **Skill development**: Technical skills often follow a pattern where fundamental abilities develop quickly, but mastery takes much longer. The logarithmic curve in Model 2 better represents this pattern.
4. **Statistical evidence**: The higher R² for Model 2 supports the economic theory that the relationship between experience and performance is non-linear.
5. **Ceiling effect**: Performance ratings have a natural ceiling (100 points in this case). The logarithmic model approaches an upper limit as experience increases, which is more realistic than the linear model that would eventually predict ratings exceeding the maximum possible score. For example, Model 1 would predict ratings above 100 for artists with more than 36 years of experience, which is implausible.

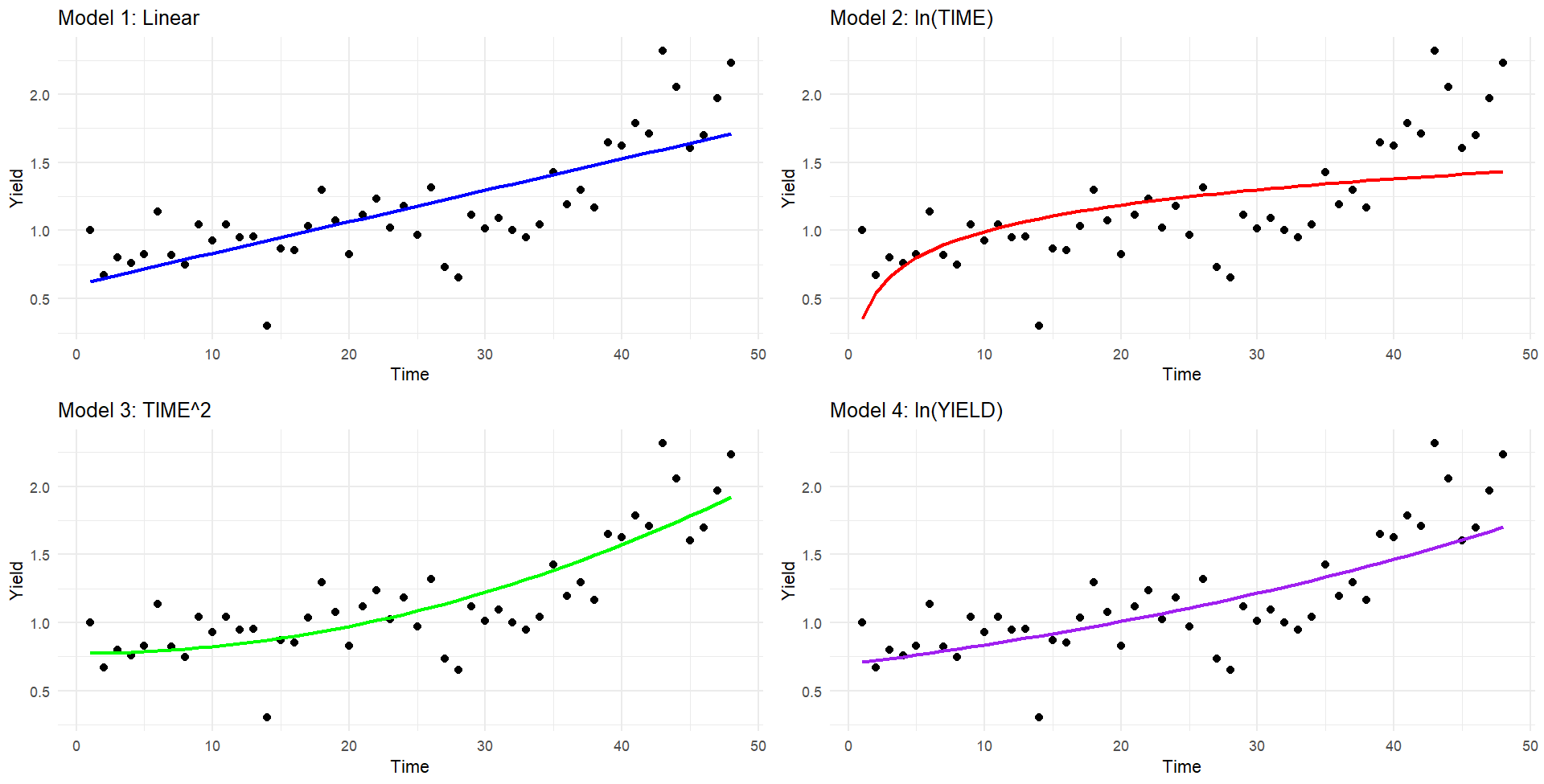
In conclusion, Model 2 not only fits the data better statistically but also aligns with economic theories of human capital development and learning curves, making it the more reasonable model.

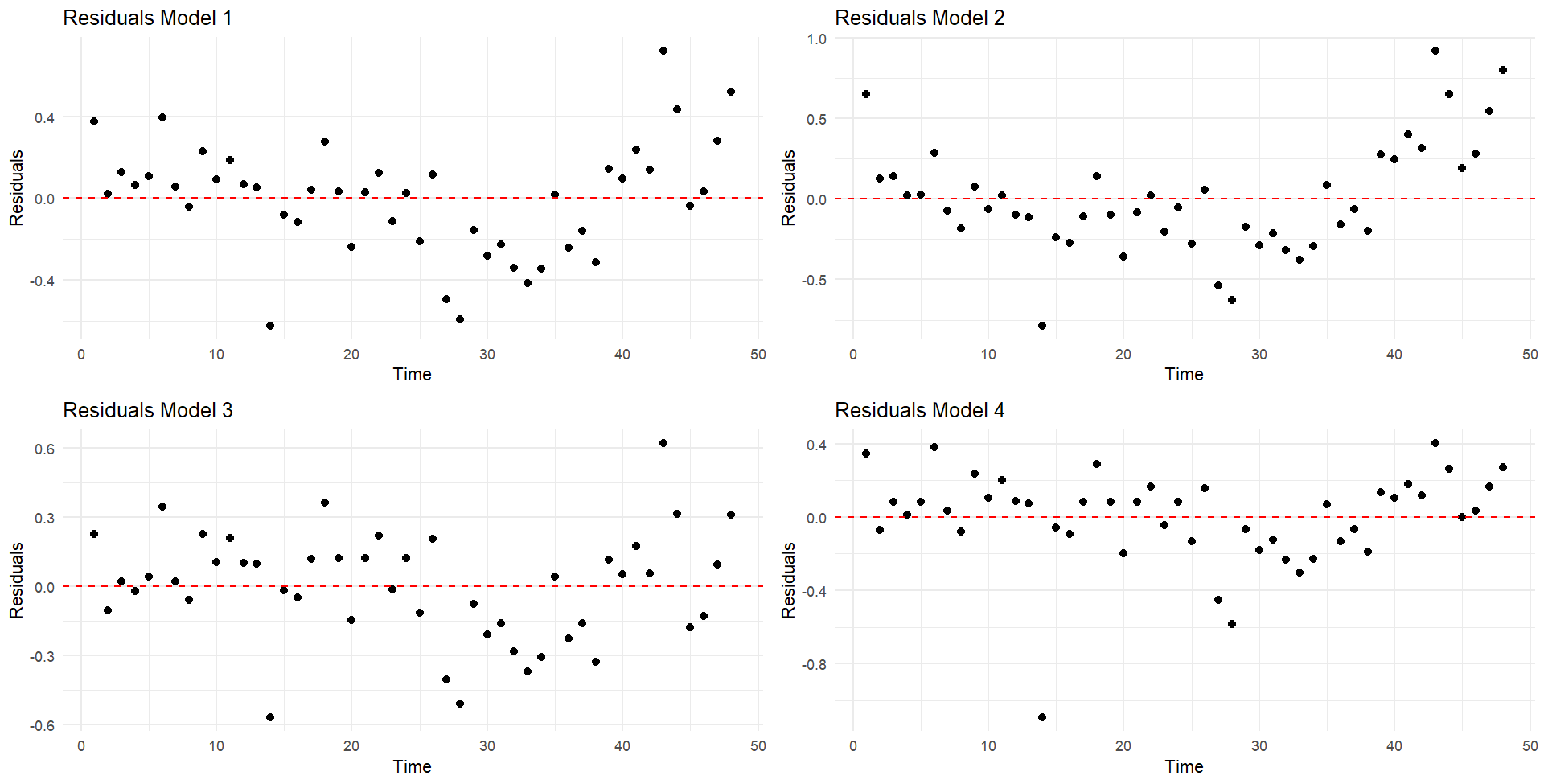


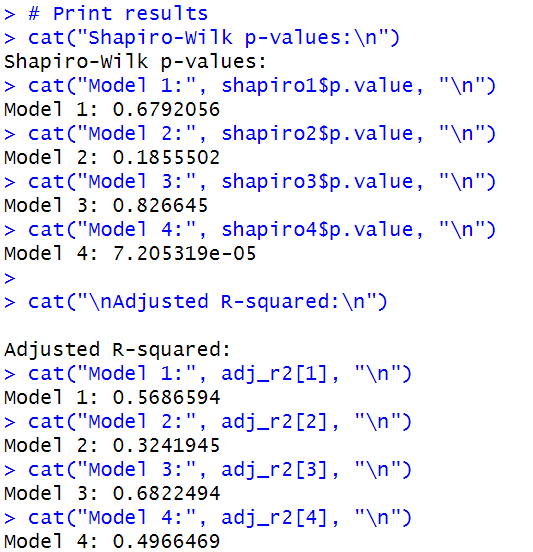


# **Question 28**

a.







**Model 3 (TIME²)** appears to be the best model:

* Best Adjusted R² → explains most variance.
* Best residual distribution → passes Shapiro-Wilk normality test (p = 0.8266).
* Visual fit and residuals also support this.

b.

A math equation with black text

AI-generated content may be incorrect.

c.

A close-up of numbers

AI-generated content may be incorrect.

d.

A screenshot of a computer code

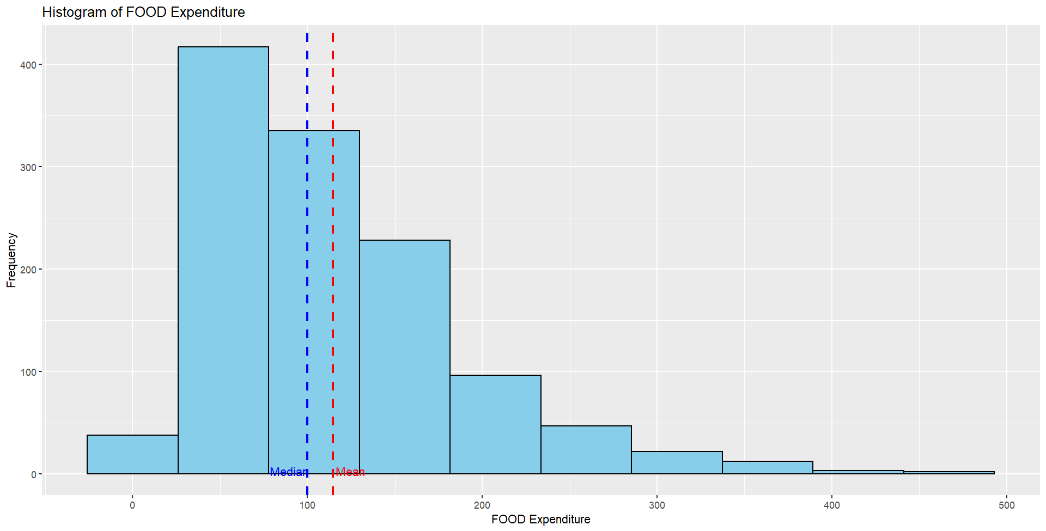
AI-generated content may be incorrect.

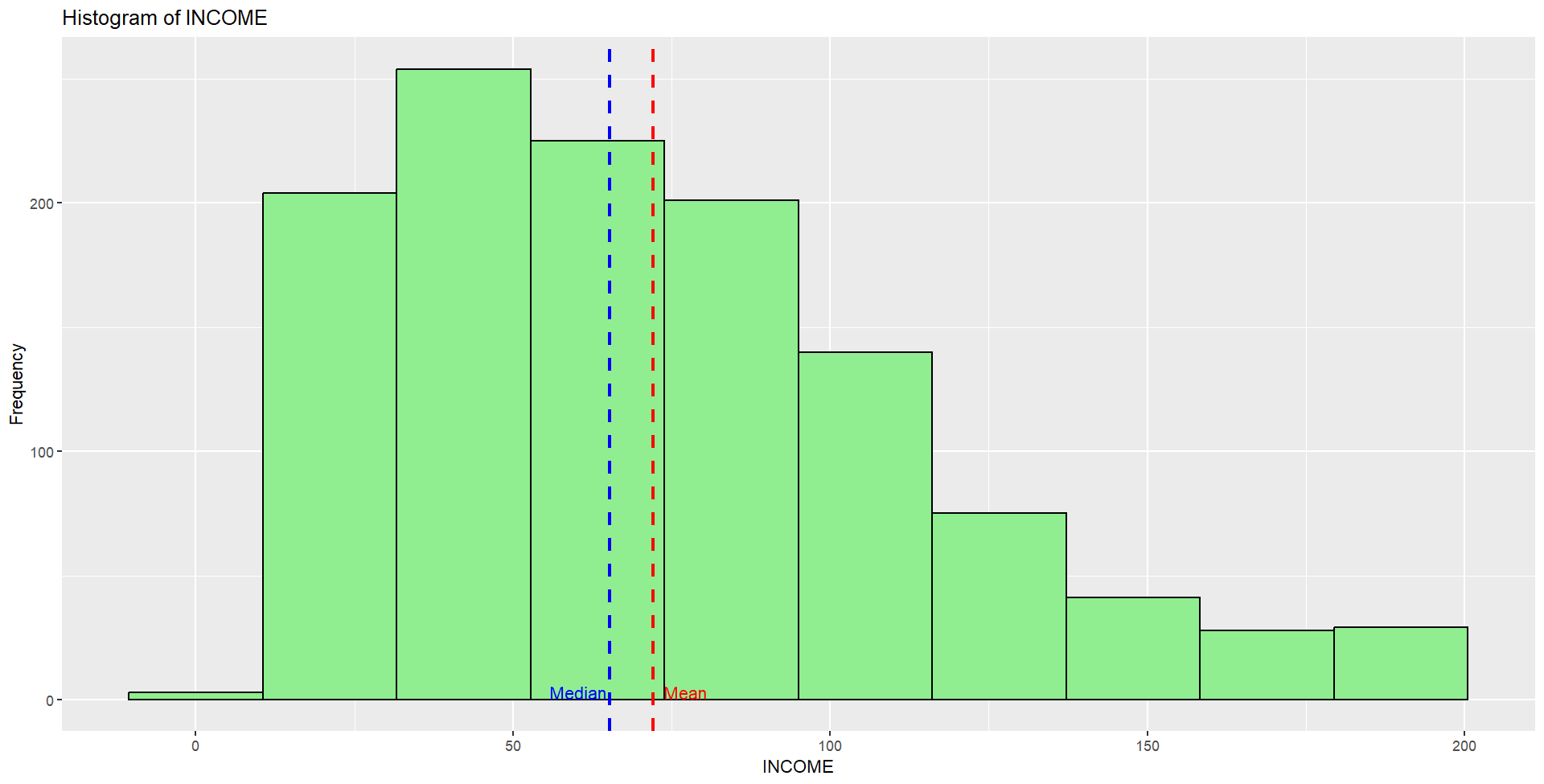
# **Question 29**

a. Summary statistics and histograms

A screenshot of a computer screen

AI-generated content may be incorrect.





A screenshot of a computer code

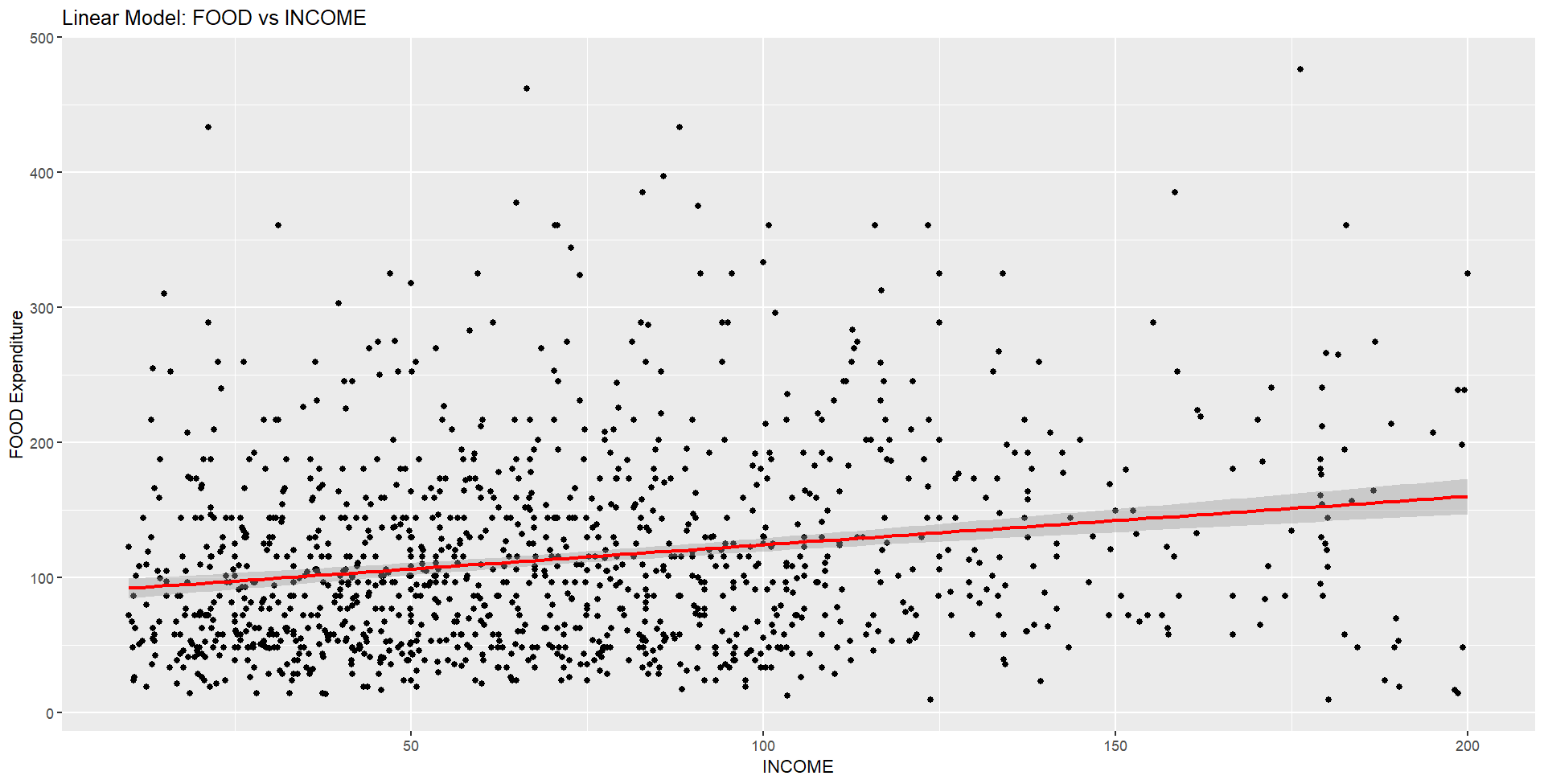
AI-generated content may be incorrect.

Both distributions are positively skewed with mean’s greater than medians. They are not bell shaped or symmetrical. For INCOME the Jarque-Bera statistic is 148.21 and for FOOD expenditure it is 648.65. The critical value for a test at the 5% level is 5.99. We reject the null hypothesis of normality for each variable.

b.

A screenshot of a computer

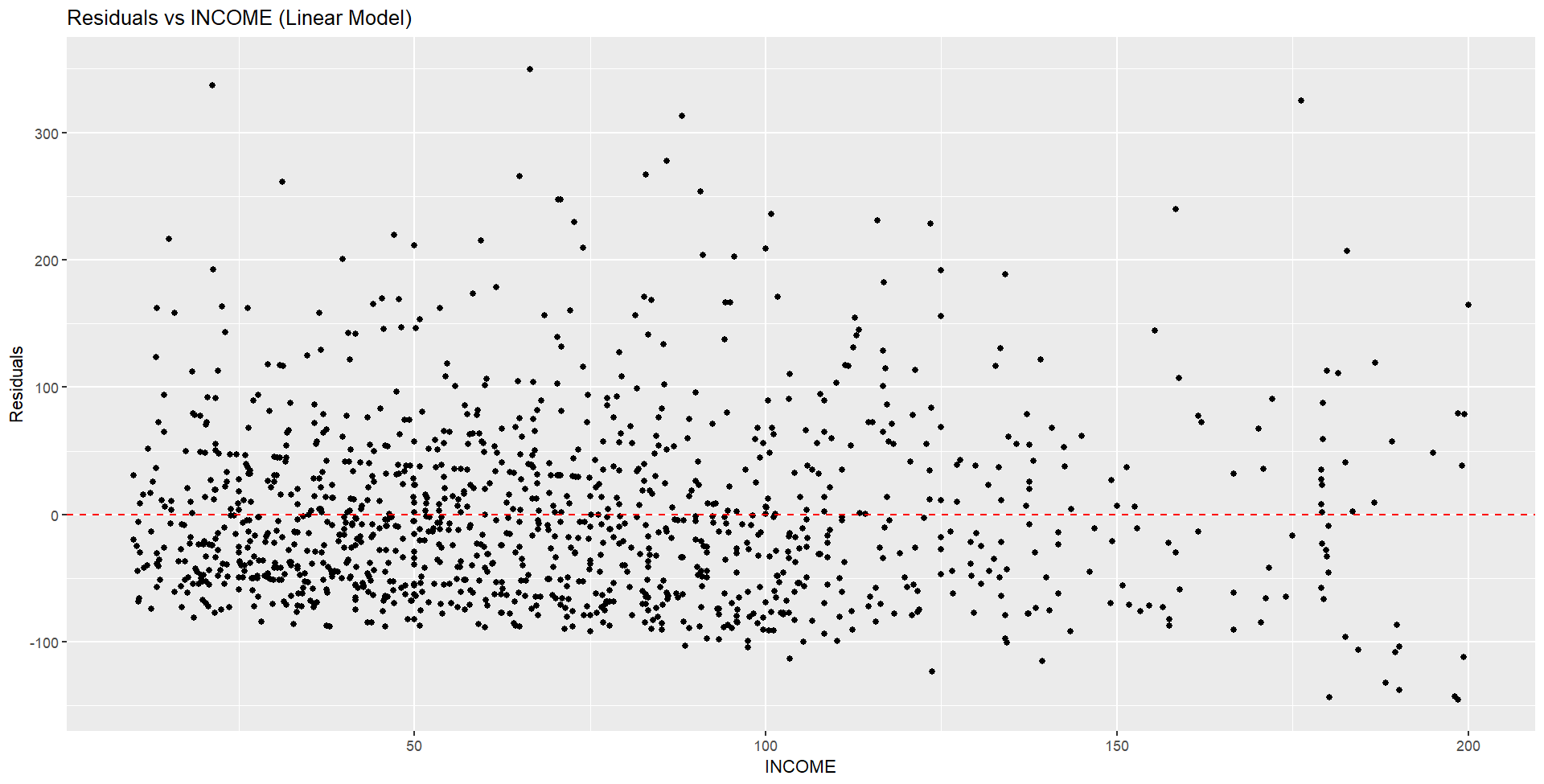
AI-generated content may be incorrect.



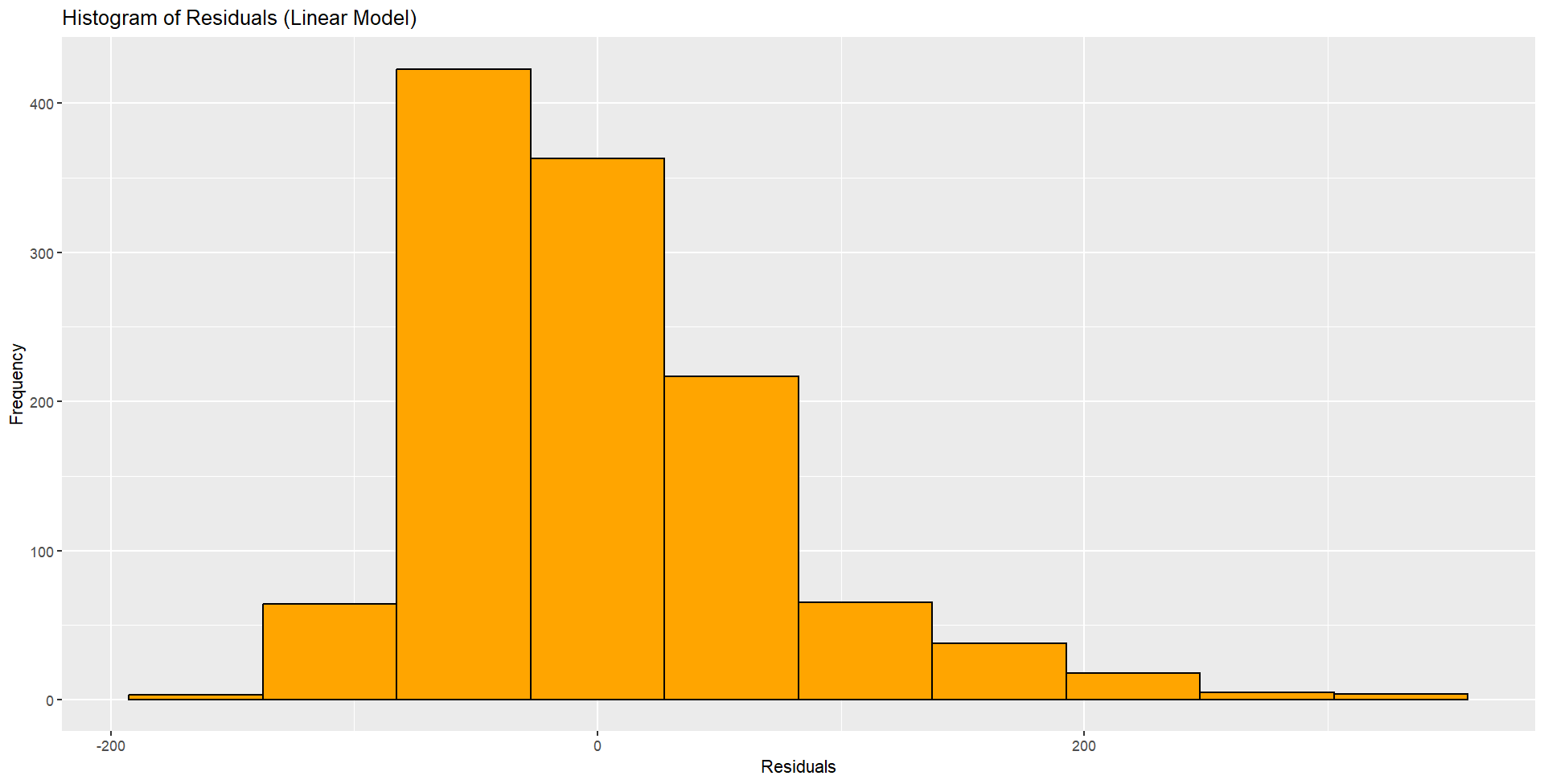
95% Confidence Interval for β₂: 0.2619215 to 0.455452

We have not estimated the effect of changing income on average food consumption very precisely: the R² is low, and the confidence interval is wide.

c.



The least squares residuals are plotted above. 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 624.186, which is far greater than the 5% critical value 5.99.



A white background with black text

AI-generated content may be incorrect.

It is more important for the random error e to be normally distributed than for the variables FOOD and INCOME to be normally distributed. This is because the statistical inference (hypothesis tests, confidence intervals) in regression analysis relies on the assumption that the error term follows a normal distribution, not the variables themselves.

d.

A screenshot of a computer

AI-generated content may be incorrect.

The estimated elasticities are **dissimilar** — they **increase noticeably** as income increases.

The interval estimates **do not overlap**, supporting the idea that the elasticities are **meaningfully different** across income levels.

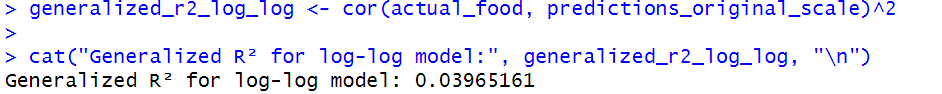
Based on economic principles, the income elasticity for food should decrease as income increases, as food is a necessity. Higher income households spend a smaller proportion of additional income on food compared to lower income households.

But in this data, it’s increasing, possibly reflecting low-to-middle income levels, where people still expand their food consumption as income rises.

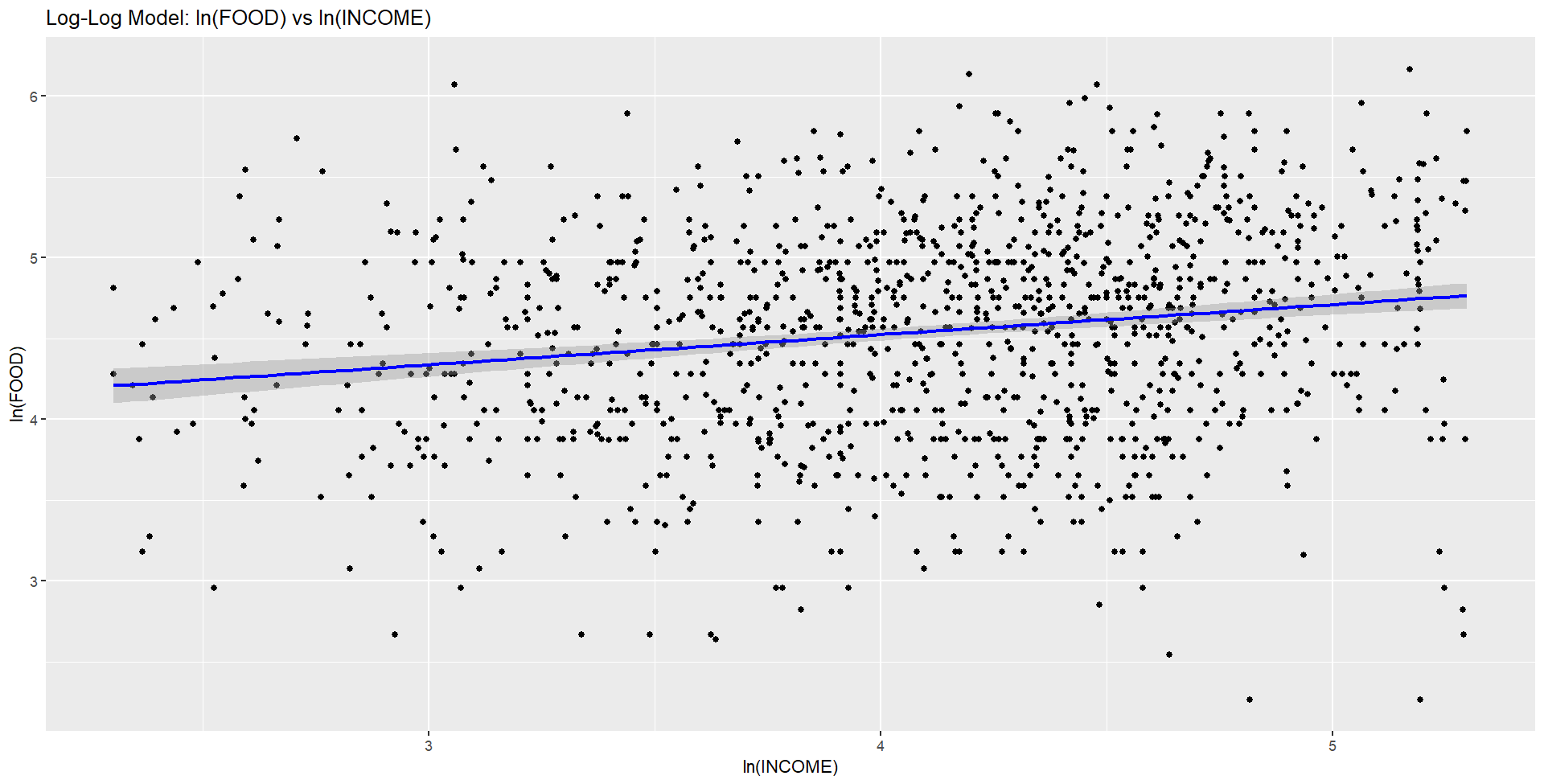
e.

A screenshot of a computer

AI-generated content may be incorrect.



The generalized R2 is 0.03965 which is slightly smaller than the R2 from the linear model (0.04228). The log-log model provides a valid functional form but does not improve model fit over the linear specification in this case. Based on both the R² comparison and visual inspection of the scatter plots, the linear model is slightly preferred.



f.

Elasticity estimate for Log-Log Model: 0.1863054

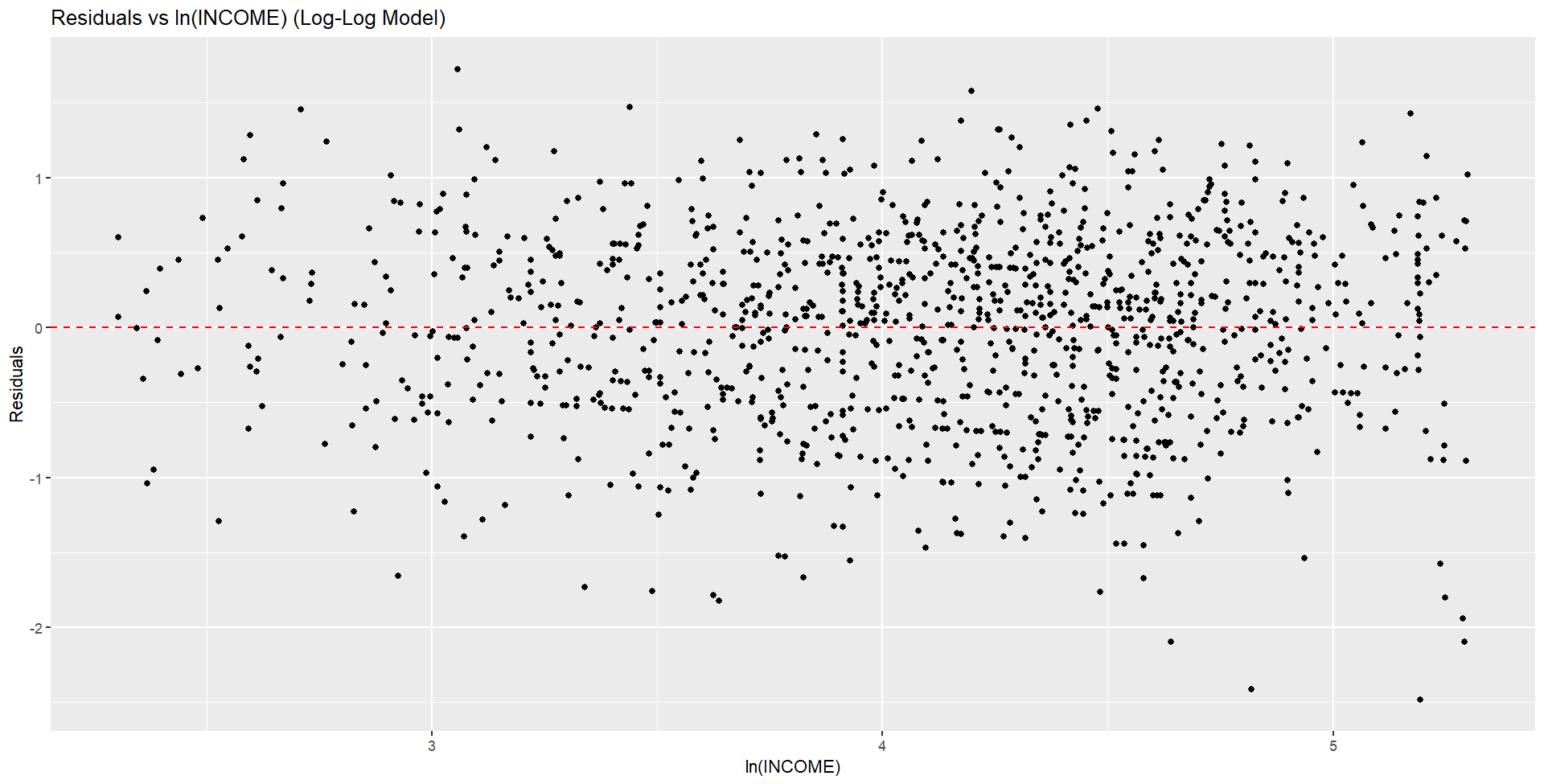
95% Confidence Interval for elasticity: 0.1293432 to 0.2432675

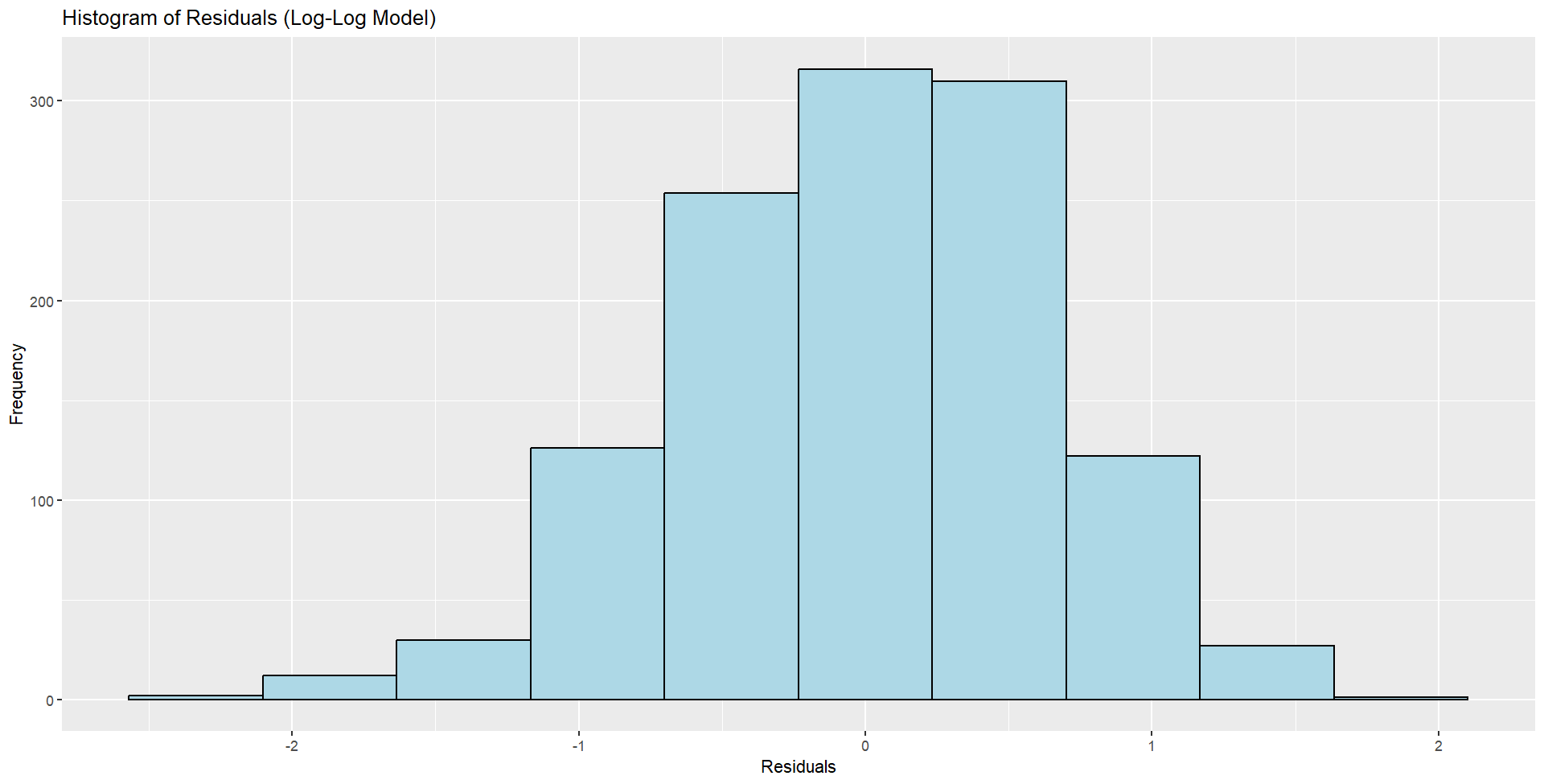
The elasticity estimate from the log-log model is **statistically similar** to that of the linear model at moderate income levels (e.g. INCOME = 65). This is supported by the **overlapping confidence intervals**, indicating no significant difference in estimated elasticity at that point.

A screenshot of a computer program

AI-generated content may be incorrect.

g.





A computer screen shot of a computer code

AI-generated content may be incorrect.

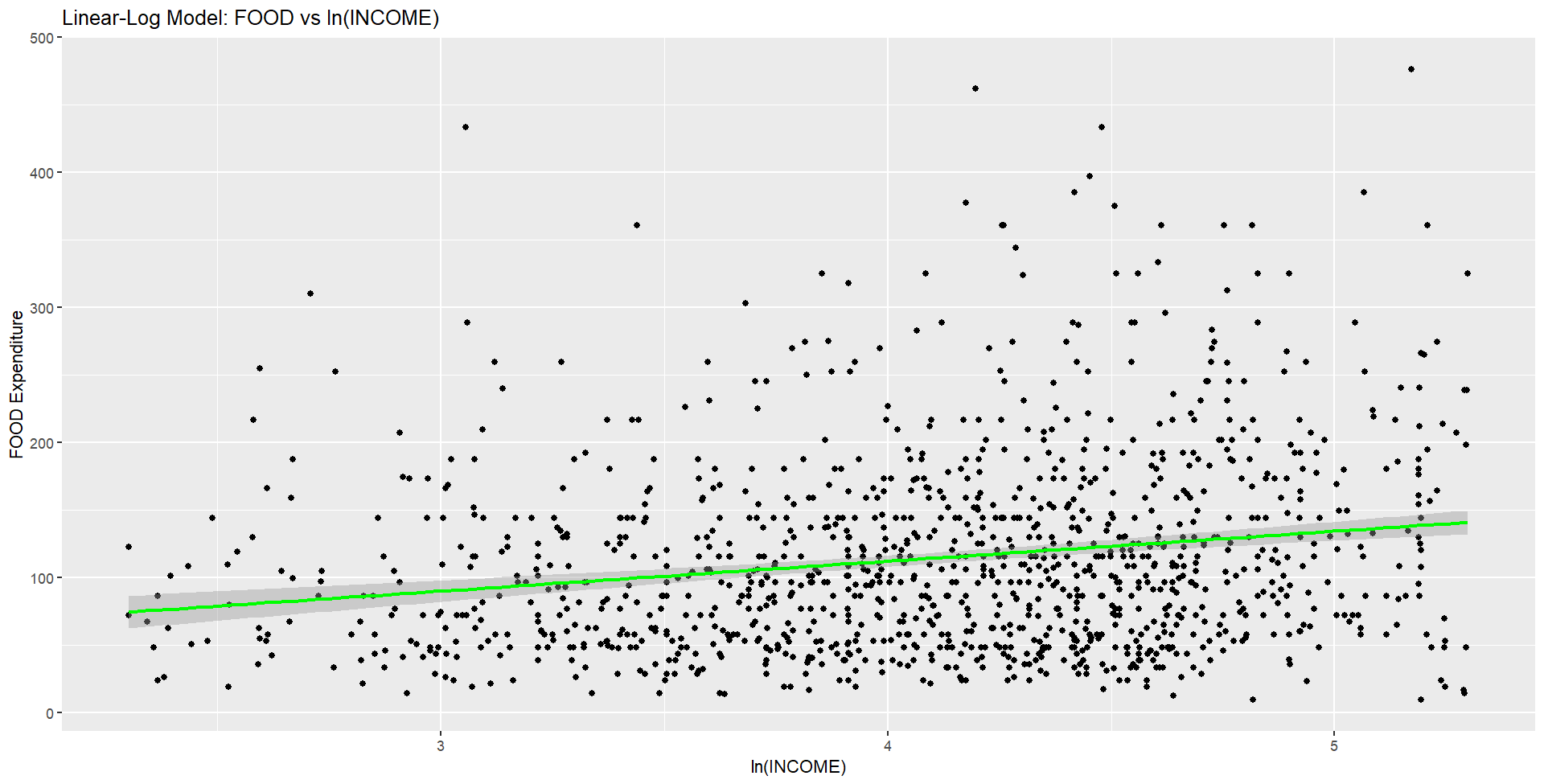
The residuals from the log-log model show no major violations of regression assumptions. The scatter appears random and symmetric, with no obvious patterns and only a slight suggestion of heteroskedasticity. This supports the choice of the **log-log model** as a statistically well-behaved specification relative to the alternatives.

There is a slight negative skew (Skewness = -0.3577) and Kurtosis is 3.0719. 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.

h.

A screenshot of a computer

AI-generated content may be incorrect.



A close-up of a code

AI-generated content may be incorrect.

The figure is similar to that of the linear model and not as well-defined as the plot for the log-log model. Based on this visual criterion, the log-log model appears to fit the data better.

The R² value for the linear-log model is 0.038, which is lower than that of the linear model and also lower than the generalized R² from the log-log model. Using this statistical criterion, the linear model seems to provide a better fit.

i.

A screenshot of a computer

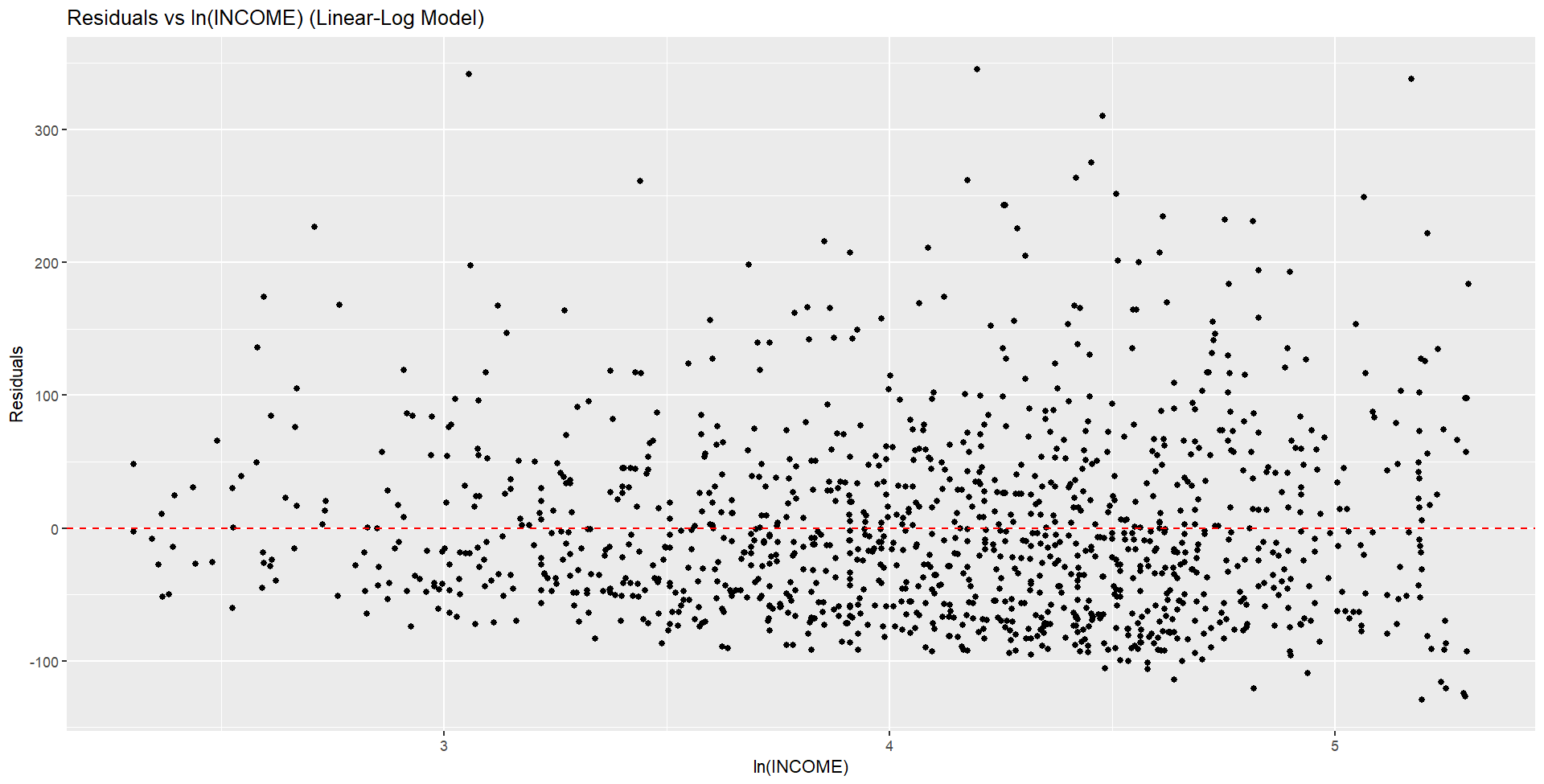
AI-generated content may be incorrect.

A computer screen shot of a number

AI-generated content may be incorrect.

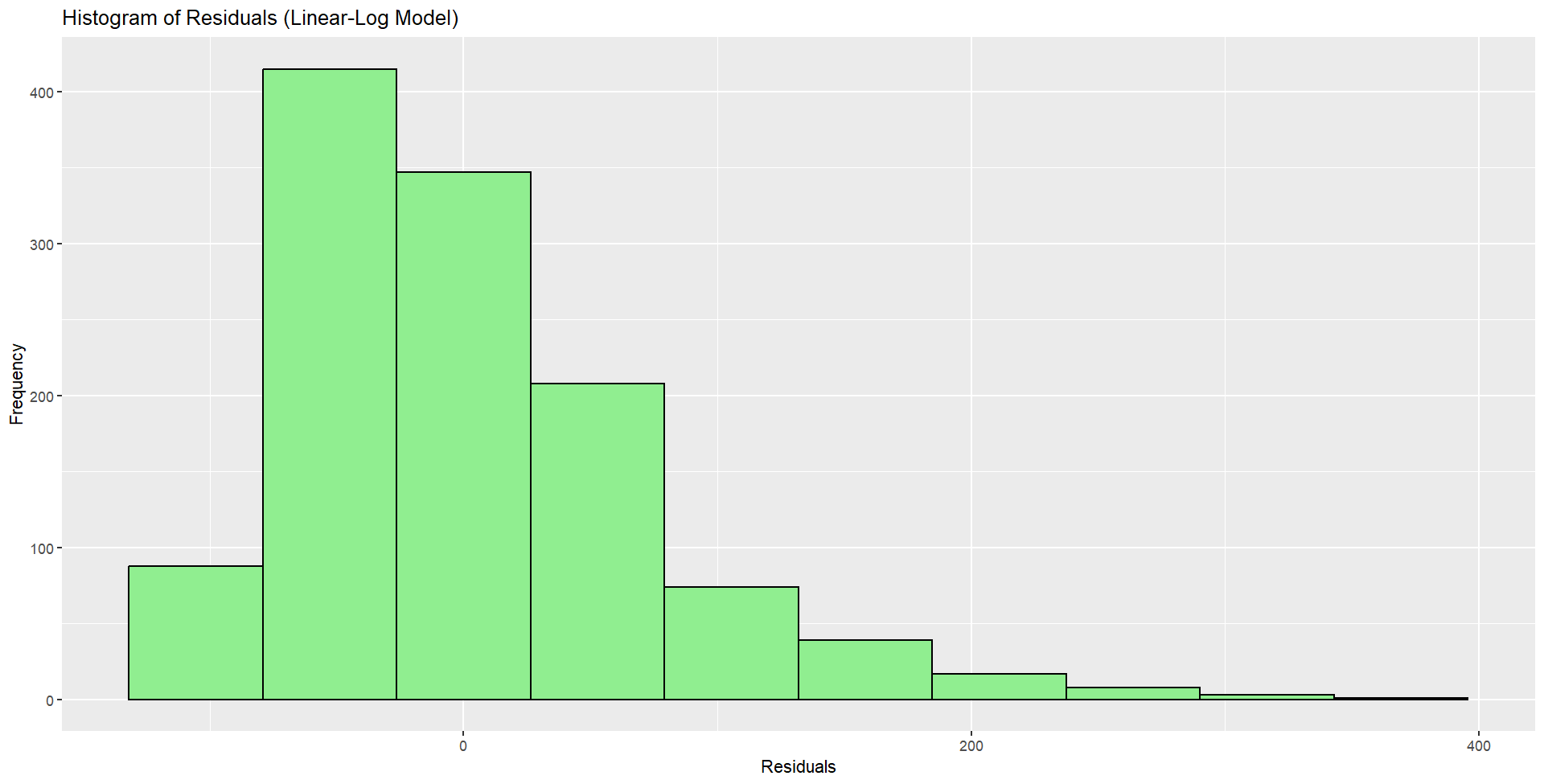
The elasticity estimate from the **log-log model** is **statistically similar** to the elasticity from the linear and linear-log models at moderate income levels (e.g. INCOME = 65), and lies within the range of linear model elasticities.  
However, the **linear-log model** shows **dissimilar elasticities** at low and high income levels when compared to the linear model, suggesting that the choice of model significantly affects elasticity estimates across different income ranges.

j.



A white background with black text

AI-generated content may be incorrect.



**How to interpret residual skewness visually:**

* **Positive skew**: More residuals **below the zero line** (many small negatives, few large positives).
* **Negative skew**: More residuals **above the zero line** (many small positives, few large negatives).
* **Symmetry**: Residuals are roughly balanced above and below zero.

The residual scatter shows positive skewness at each income level and overall. 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.

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.