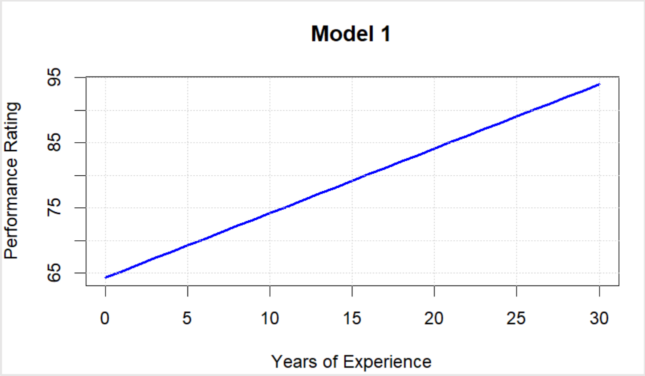
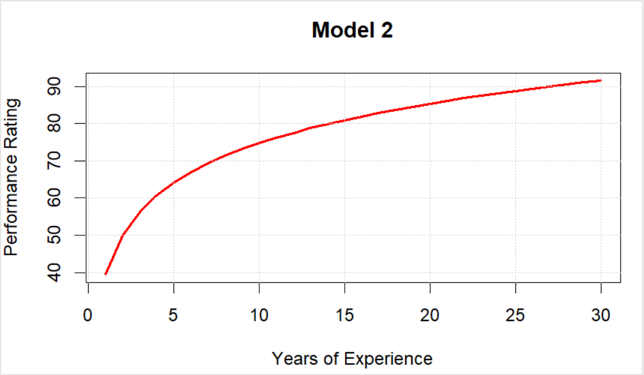
4.4

(a)

(b) Since ln(0) does not exist, four artists with no experience are excluded.

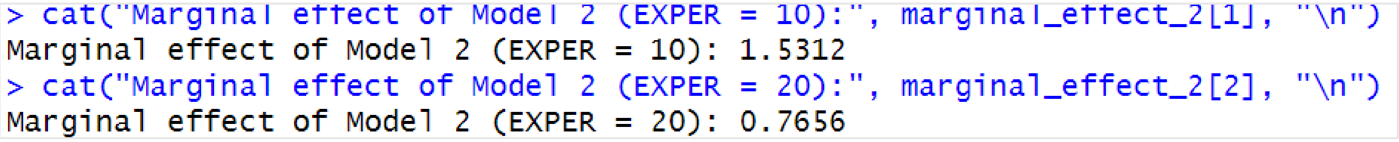




(c)

For linear regression model, the slope (β_2) of Model 1 is 0.99, which means that regardless of the amount of experience, each additional year of experience increases the RATING by 0.990.

(d)



```
> cat("Marginal effect of Model 2 (EXPER = 10):", marginal_effect_2[1], "\n")
Marginal effect of Model 2 (EXPER = 10): 1.5312
> cat("Marginal effect of Model 2 (EXPER = 20):", marginal_effect_2[2], "\n")
Marginal effect of Model 2 (EXPER = 20): 0.7656
```

As experience increases, the impact of each additional year of experience decreases, which aligns with the principle of diminishing returns.

(e)

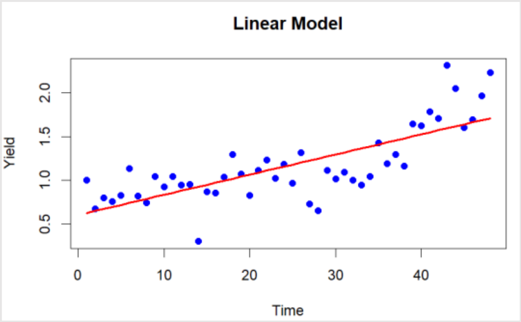
Since Model 2 has the highest R^2value (0.6414), it provides a better fit for the data.

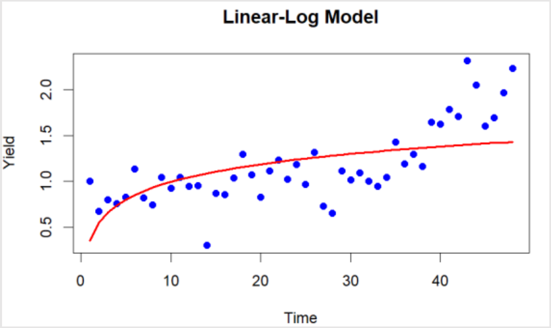
(f)

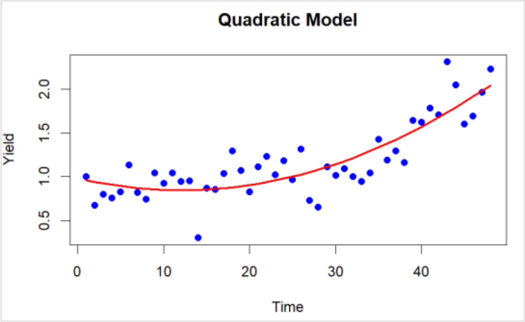
Model 2 is more reasonable because its marginal effect of technical artists should be diminishing; rapid growth in the early stages, stabilizing later. But Model 1 assumes the impact of experience is constant, which does not align with reality.

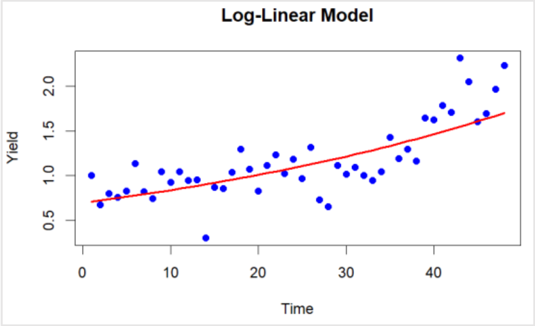
4.28

(a)(i)

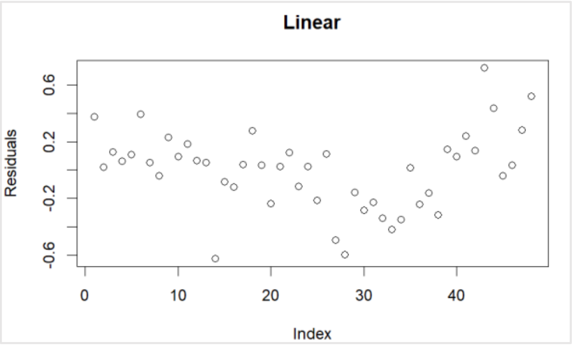


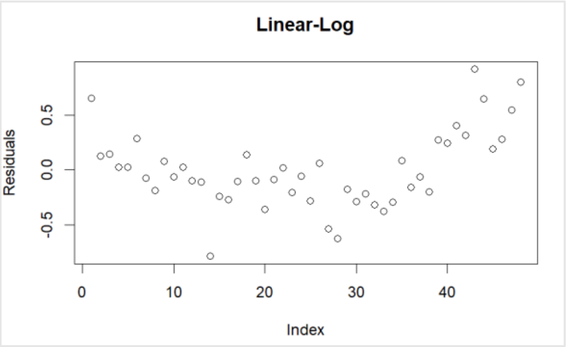


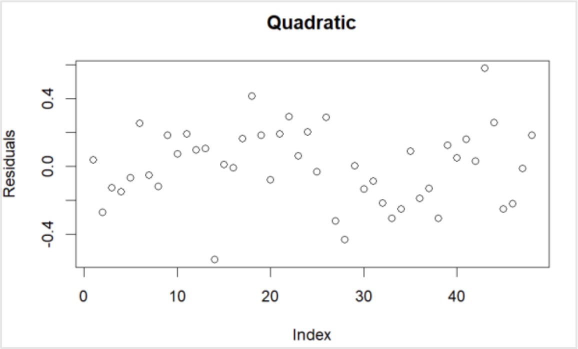


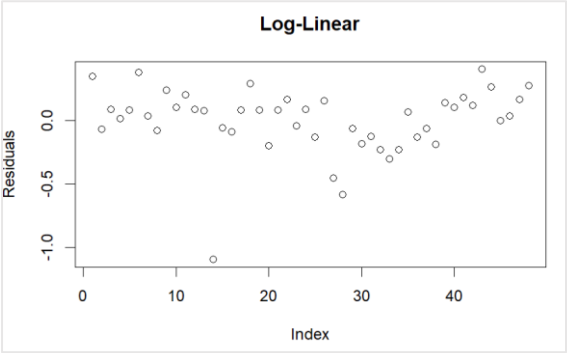


(ii)

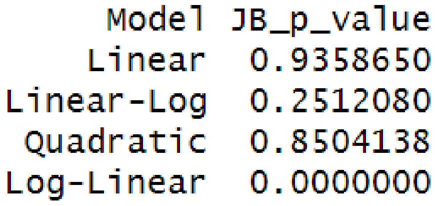








(iii)error normality test



```
      Model  JB_p_value
     Linear   0.9358650
 Linear-Log   0.2512080
  Quadratic   0.8504138
 Log-Linear   0.0000000
```

If P-value > 0.05, then we can reject H_0. The linear model and the quadratic model may have normal distribution, but the quadratic model has the greatest P-value, so we can say that better conforms to the normality assumption.

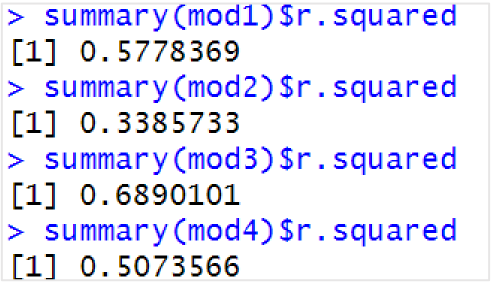
(iv) R^2

The quadratic model has the greatest R² value, indicating the best fit among the four models.

Based on the evaluation criteria:

(i) The quadratic model provides a better visual fit to the data.

(ii) The residuals of the quadratic and linear model show less systematic pattern, indicating a better fit.



```
> summary(mod1)$r.squared
[1] 0.5778369
> summary(mod2)$r.squared
[1] 0.3385733
> summary(mod3)$r.squared
[1] 0.6890101
> summary(mod4)$r.squared
[1] 0.5073566
```

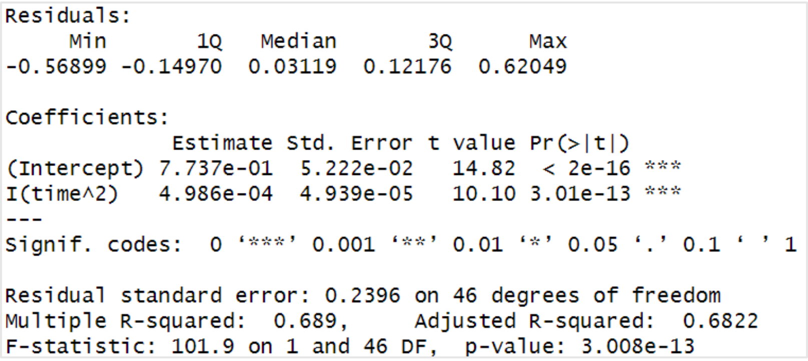
(iii)The JB test suggests that the residuals of the quadratic model conform more closely to normality.

(iv)The quadratic model has the highest R², indicating the best explanatory power.

Conclusion:

The quadratic model is the preferred specification as it provides the best overall fit, has more normally distributed residuals, and explains the variation in wheat yield more effectively.

(b)



```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.56899 -0.14970  0.03119  0.12176  0.62049 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.737e-01  5.222e-02   14.82  < 2e-16 ***
I(time^2)   4.986e-04  4.939e-05   10.10 3.01e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2396 on 46 degrees of freedom
Multiple R-squared:  0.689,	Adjusted R-squared:  0.6822 
F-statistic: 101.9 on 1 and 46 DF,  p-value: 3.008e-13
```

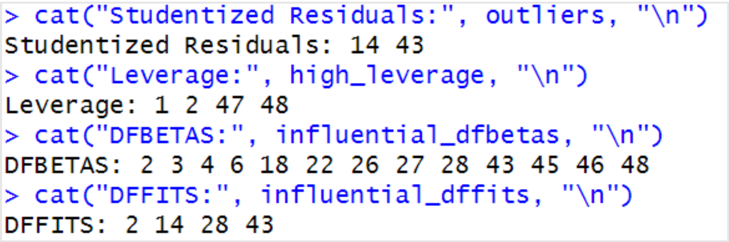
-The coefficient of TIME2 is 0.0004986 and is a positive number, indicating that YIELD grows at an accelerated rate over time rather than increasing linearly.

-The value of the TIME2 coefficient = 10.10, indicating that the variable is extremely significant in the regression.

-P-value = 3.01e-13, which is much smaller than 0.05, indicating that at a 99.9% confidence level, the null hypothesis that the TIME2 coefficient is 0 can be rejected.

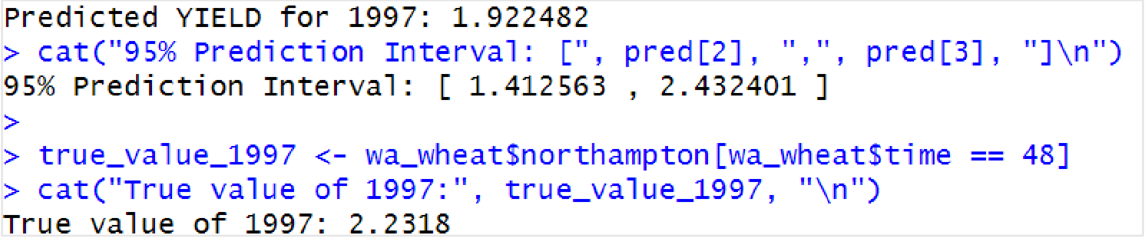
-The TIME2 variable is extremely significant in this model, indicating that it has a very strong impact on YIELD.

(c)



```
> cat("Studentized Residuals:", outliers, "\n")
Studentized Residuals: 14 43 
> cat("Leverage:", high_leverage, "\n")
Leverage: 1 2 47 48 
> cat("DFBETAS:", influential_dfbetas, "\n")
DFBETAS: 2 3 4 6 18 22 26 27 28 43 45 46 48 
> cat("DFFITS:", influential_dffits, "\n")
DFFITS: 2 14 28 43 
```

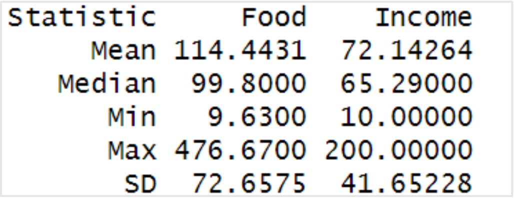
(d) Yes, the 95 % prediction interval contains the true value of 1997.



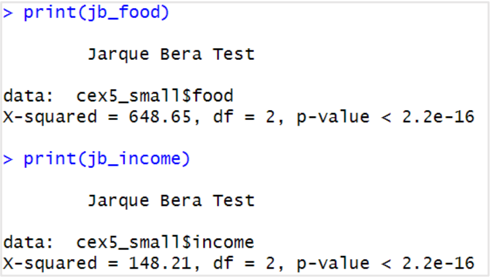
```
Predicted YIELD for 1997: 1.922482 
> cat("95% Prediction Interval: [", pred[2], ",", pred[3], "]\n")
95% Prediction Interval: [ 1.412563 , 2.432401 ]
> 
> true_value_1997 <- wa_wheat$northampton[wa_wheat$time == 48]
> cat("True value of 1997:", true_value_1997, "\n")
True value of 1997: 2.2318 
```

4.28

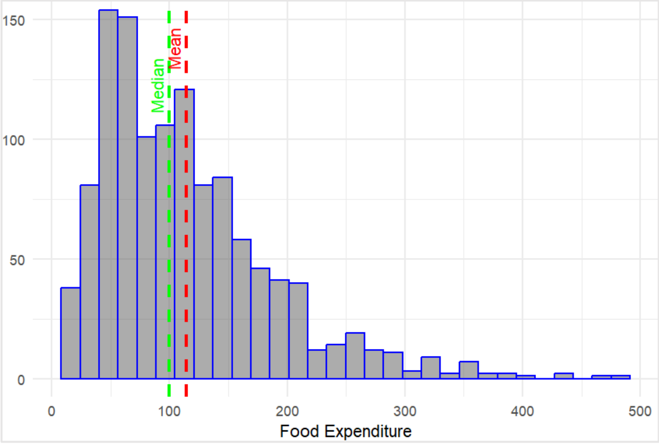
(a)

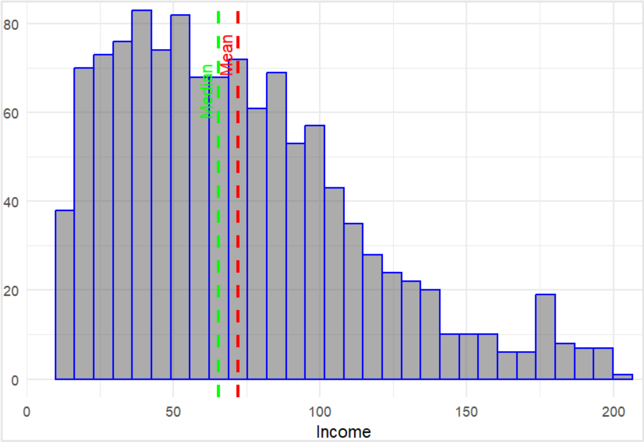


```
Statistic     Food    Income
     Mean 114.4431  72.14264
   Median  99.8000  65.29000
      Min   9.6300  10.00000
      Max 476.6700 200.00000
       SD  72.6575  41.65228
```



```
> print(jb_food)

	Jarque Bera Test

data:  cex5_small$food
X-squared = 648.65, df = 2, p-value < 2.2e-16

> print(jb_income)

	Jarque Bera Test

data:  cex5_small$income
X-squared = 148.21, df = 2, p-value < 2.2e-16
```

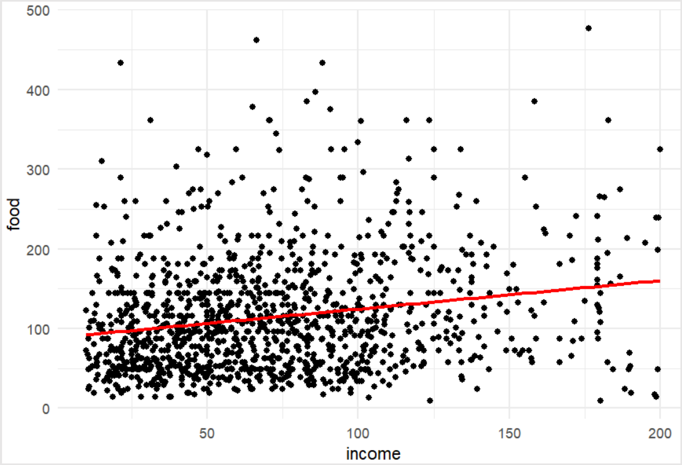


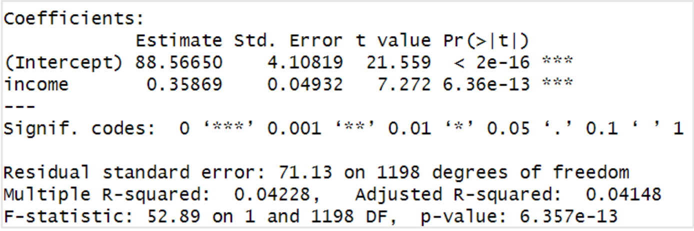


Both distributions are positively skewed with mean’s greater than medians, they are not bell shaped or symmetrical. Since the p-value < 0.05, we reject the null hypothesis of normality for each variable.

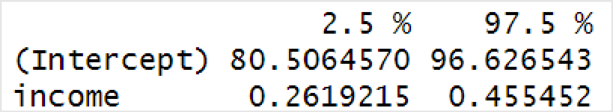
(b)

FOOD = 88.5665 + 0.3587 INCOME + e





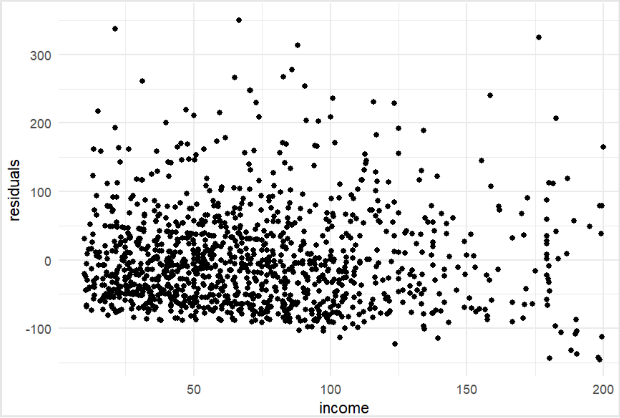
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 88.56650    4.10819  21.559  < 2e-16 ***
income       0.35869    0.04932   7.272 6.36e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 71.13 on 1198 degrees of freedom
Multiple R-squared:  0.04228,	Adjusted R-squared:  0.04148 
F-statistic: 52.89 on 1 and 1198 DF,  p-value: 6.357e-13
```

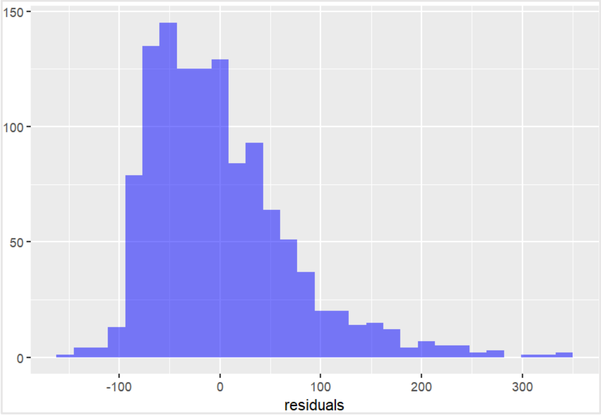


```
                 2.5 %    97.5 %
(Intercept) 80.5064570 96.626543
income       0.2619215  0.455452
```

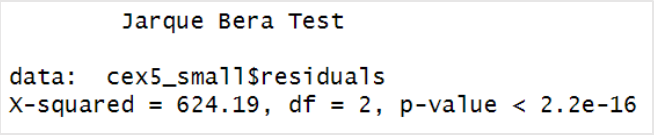
R^2 = 0.04228, low explanatory power indicates that income is not the primary factor influencing food expenditure.

(c)





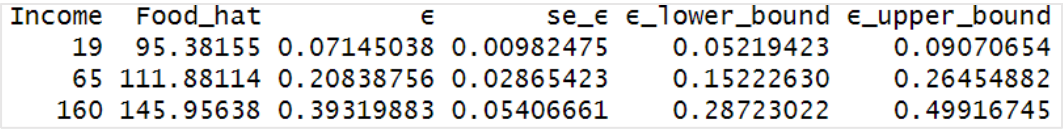
The positive skew at each income is clear. There is not a clear “spray” pattern except at high incomes. The residual histogram shows the skewness.



```
	Jarque Bera Test

data:  cex5_small$residuals
X-squared = 624.19, df = 2, p-value < 2.2e-16
```

The p-value of Jarque-Bera test <0.05, so we reject null hypothesis. It means that the residuals of linear model may not follow the normal distribution.

(d)



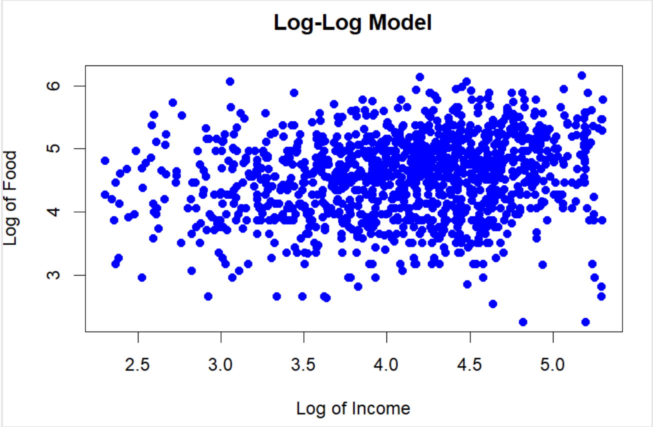
```
Income   Food_hat          ϵ        se_ϵ ϵ_lower_bound ϵ_upper_bound
    19  95.38155 0.07145038 0.00982475    0.05219423    0.09070654
    65 111.88114 0.20838756 0.02865423    0.15222630    0.26454882
   160 145.95638 0.39319883 0.05406661    0.28723022    0.49916745
```

- Elasticity increases with income, indicating that the income elasticity of food expenditure varies significantly across different income levels.

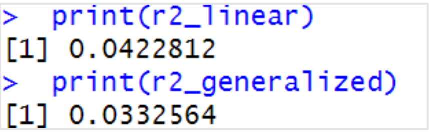
- The confidence intervals for elasticity do not overlap, suggesting that food expenditure behavior differs significantly among income groups from a statistical perspective.

- Food is considered a necessity, and income elasticity of food expenditure is generally expected to decrease or stabilize as income increases. However, the results show that elasticity increases with income, which is not entirely consistent with standard economic predictions.

(e)

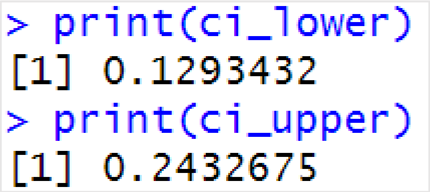


The generalized R^2 is 0.03326 which is slightly smaller than the R^2 from the linear model.

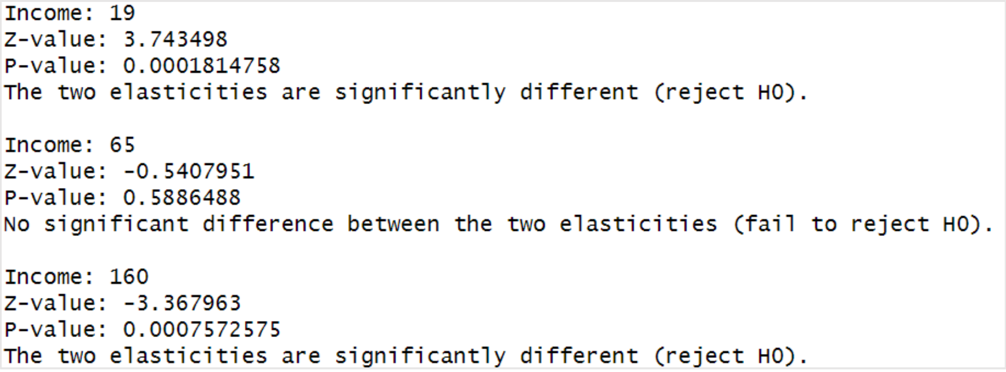


```
> print(r2_linear)
[1] 0.0422812
> print(r2_generalized)
[1] 0.0332564
```

(f)

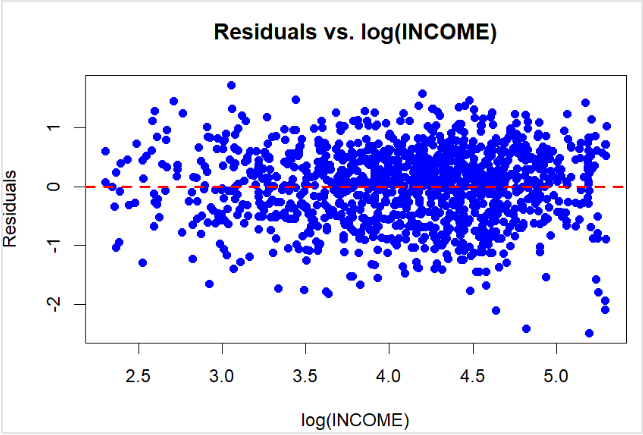


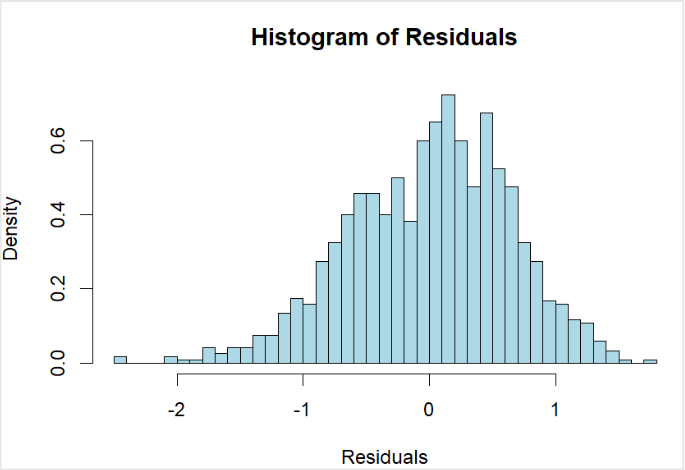
```
> print(ci_lower)
[1] 0.1293432
> print(ci_upper)
[1] 0.2432675
```

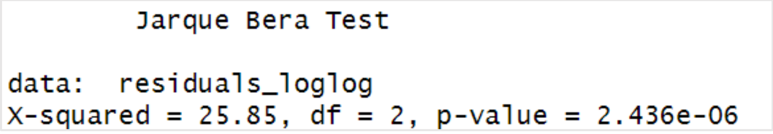


```
Income: 19
Z-value: 3.743498
P-value: 0.0001814758
The two elasticities are significantly different (reject H0).

Income: 65
Z-value: -0.5407951
P-value: 0.5886488
No significant difference between the two elasticities (fail to reject H0).

Income: 160
Z-value: -3.367963
P-value: 0.0007572575
The two elasticities are significantly different (reject H0).
```

(g)



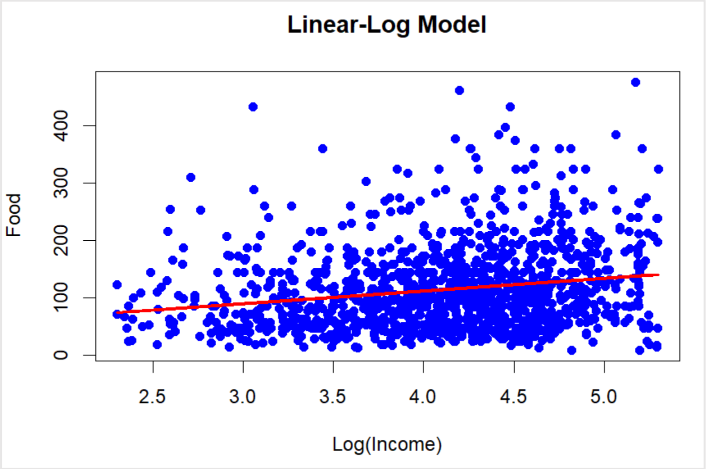


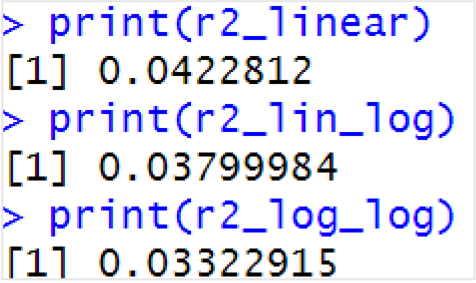


```
        Jarque Bera Test

data:  residuals_loglog
X-squared = 25.85, df = 2, p-value = 2.436e-06
```

The JarqueBera 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)

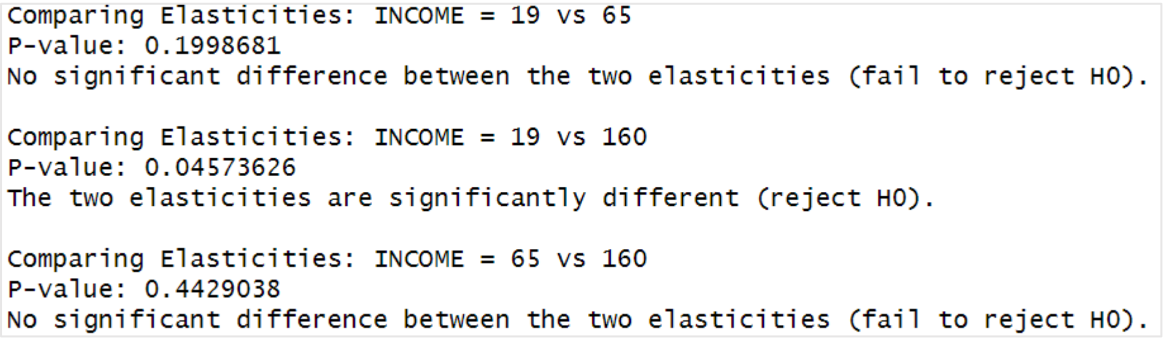




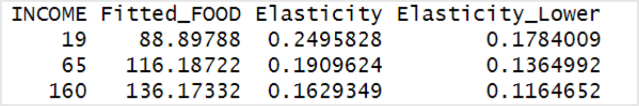
```
> print(r2_linear)
[1] 0.0422812
> print(r2_lin_log)
[1] 0.03799984
> print(r2_log_log)
[1] 0.03322915
```

R^2 of linear-log model is greater than the one of log-log model, but smaller than linear model. By R^2, linear model seems to fit the data better.

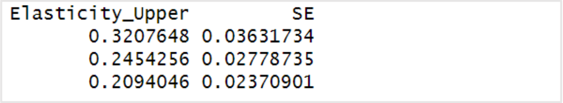
(i)



```
Comparing Elasticities: INCOME = 19 vs 65
P-value: 0.1998681
No significant difference between the two elasticities (fail to reject H0).

Comparing Elasticities: INCOME = 19 vs 160
P-value: 0.04573626
The two elasticities are significantly different (reject H0).

Comparing Elasticities: INCOME = 65 vs 160
P-value: 0.4429038
No significant difference between the two elasticities (fail to reject H0).
```

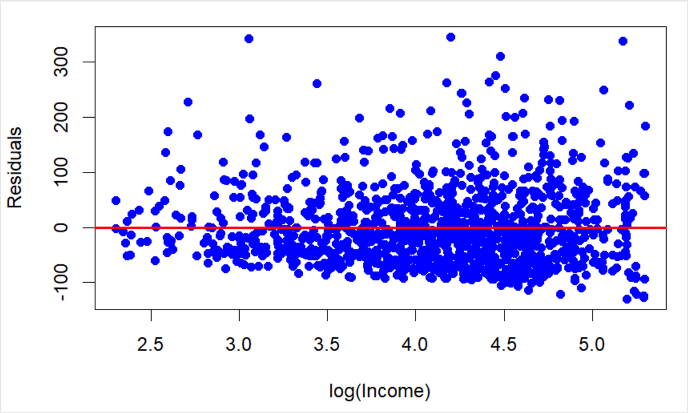


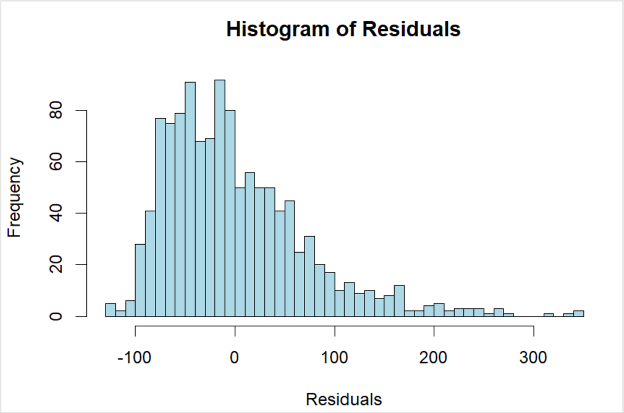
```
 INCOME  Fitted_FOOD  Elasticity  Elasticity_Lower
     19     88.89788   0.2495828         0.1784009
     65    116.18722   0.1909624         0.1364992
    160    136.17332   0.1629349         0.1164652
```

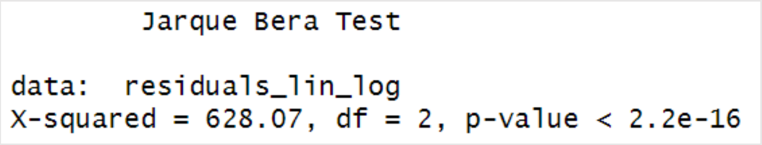


```
Elasticity_Upper          SE
       0.3207648  0.03631734
       0.2454256  0.02778735
       0.2094046  0.02370901
```

(j)







```
        Jarque Bera Test

data:  residuals_lin_log
X-squared = 628.07, df = 2, p-value < 2.2e-16
```

The JarqueBera statistic is 25.85 which is greater than the 5% critical value 5.99. So, we reject the null hypothesis that the linear-log regression errors are normal. Residuals shows positive skewness.

The data scatter suggests a slight “spray” pattern.

(k)

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, I prefer log-log model seems like a good choice.