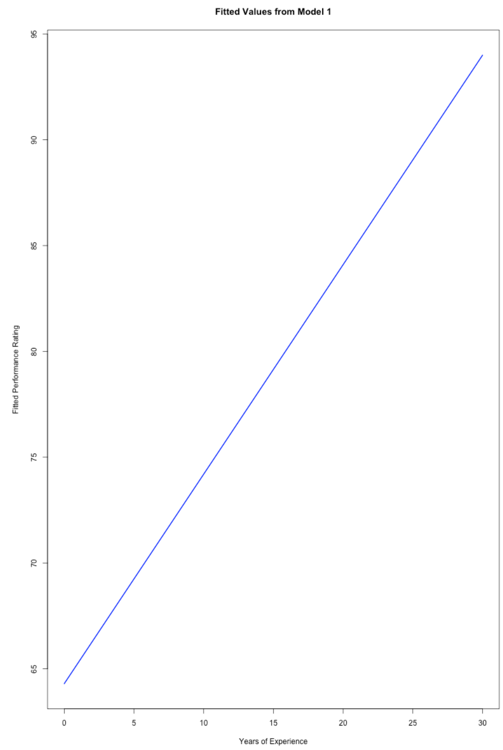
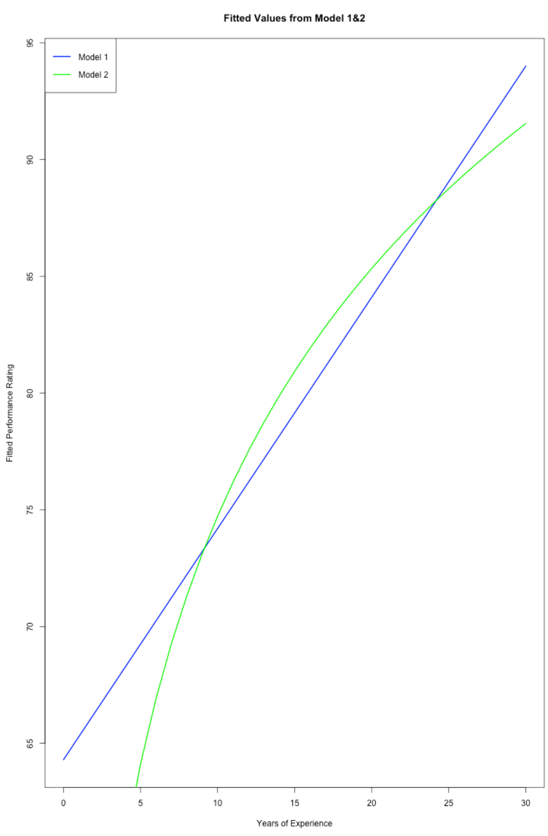
Financial econometrics: HW0317

Question: 4.4

Part a.



Part b.



In the second model the independent variable is transformed by taking it's natural logarithm. As the natural logarithm of zero is undefined, any observations where EXPER = 0 cannot be included in the transformation. So artists who have no experience (so their EXPER = 0) cannot be used in the estimate of Model 2.

Part c.

The marginal effect of EXPER on RATING can be found using the following formula as model 1 is a linear model:

$$\frac{\partial \widehat{RATING}}{\partial \text{EXPER}} = 0.990$$

The slope in this model is linear the marginal effect in both cases will be 0.990.

Part d.

The model is not linear the marginal effect of an additional year of experience is also not constant, so to compute the marginal effect we have to use the following formula:

$$\frac{\partial \widehat{RATING}}{\partial \text{EXPER}} = \frac{15.312}{\text{EXPER}}$$

Using R we find that the marginal effect of 1 year of additional experience for an artist with 10 years of experience is 1.5312 while for an artist with 20 years of experience the marginal effect is 0.7656 as can be seen here:

```
> effect_10 <- marginal_effect_model2(10)
> effect_20 <- marginal_effect_model2(20)
> effect_10
[1] 1.5312
> effect_20
[1] 0.7656
>
```

Part e.

If we look at purely the R^2 values I think that model 2 (the log model) fits the data considerably better as the R^2 for the log model are higher than the R^2 for the linear model. Whole model 1 does show that there is a relationship between experience and rating it does not capture the diminishing returns of the marginal effect which experience has on rating.

Part f.

I think model 2 is more plausible, because I think that the first 10-20 years have a higher impact on artist ratings. Purely from a standpoint that a one really negative or one really positive rating will have a less and less significant impact on the overall rating as the artists works longer and receives more ratings. I also think that the workflow of an artist in 10-20 years has changed dramatically - with new and more innovative tools being introduced constantly. So an older artist may have less exposure to those tools then someone who is younger.

Question: 4.29

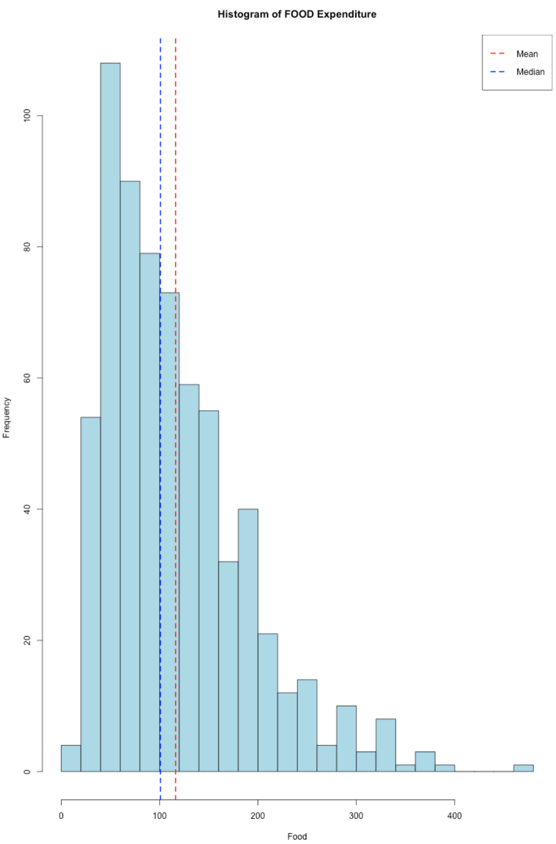
Part a.

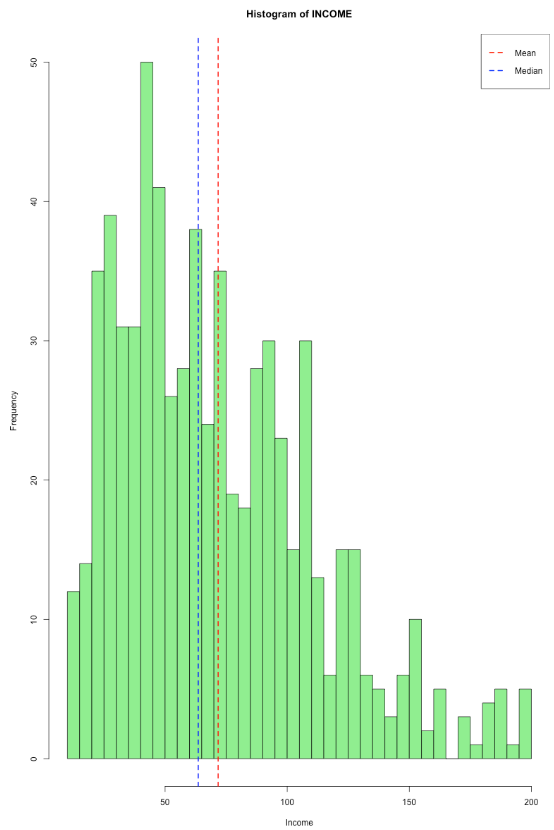
Reporting on mean, median, minimum, maximum and standard deviation of the data set for a family of 3 for food:

Item	Number
Mean	116.2501
Median	100.835
Min	14.44
Max	462.22
Standard deviation	71.16488

Reporting on mean, median, minimum, maximum and standard deviation of the data set for a family of 3 for income:

Item	Number
Mean	71.71662
Median	63.575
Min	10.83
Max	200
Standard deviation	40.28273





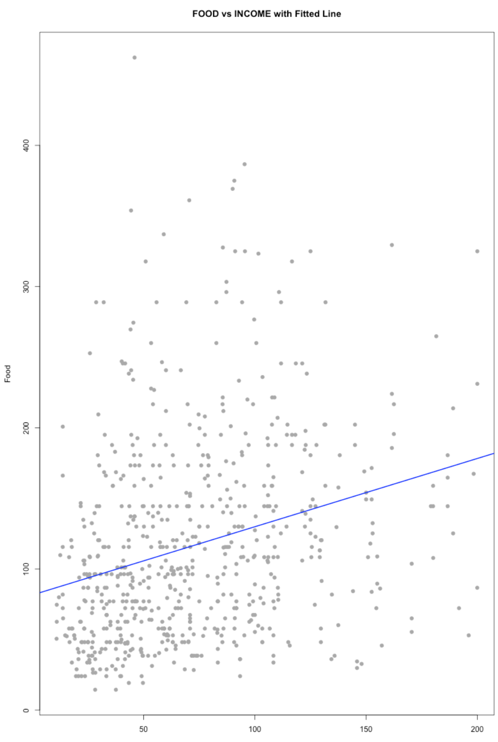
Both histograms are right-skewed rather than being bell-shaped, which indicates that the distributions are not symmetrical. In both FOOD and INCOME the the sample mean is greater than the median meaning that there are some significant outliers which increase the averages. FOOD expenditure is more heavily skewed than INCOME, with a majority of households spending lower amounts with a longer right tail. Neither of these seem to be normally distributed. In terms of the Jarque-Bera test these were the results:

```
> jarque.bera.test(df$food)

        Jarque Bera Test

data:  df$food
X-squared = 261.67, df = 2, p-value < 2.2e-16

> jarque.bera.test(df$income)

        Jarque Bera Test

data:  df$income
X-squared = 85.654, df = 2, p-value < 2.2e-16
```

This test reveals a couple of things, first of all we ca reject the null-hypothesis for both data sets as the p-values are extremely low. This test strengthens our original hypothesis that both data sets are not normally distributed.

Part b.

Below you can find the scatter plot with a fitted fitted least squares line alongside the estimated linear regression:



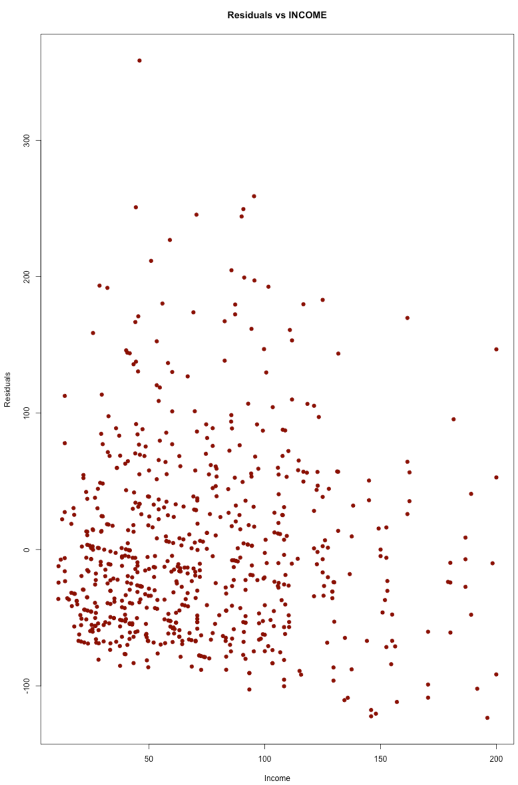

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 81.58509    5.39884  15.112  < 2e-16 ***
income       0.48336    0.06565   7.363 5.29e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 68.5 on 670 degrees of freedom
Multiple R-squared:  0.07486,	Adjusted R-squared:  0.07348
F-statistic: 54.21 on 1 and 670 DF,  p-value: 5.29e-13
```

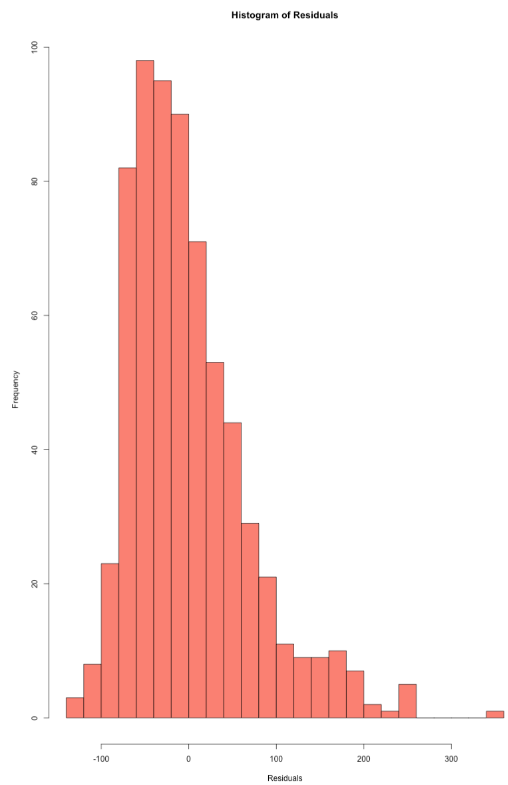
```
> confint(model_food_income, level = 0.95)
                 2.5 %     97.5 %
(Intercept) 70.9844030 92.1857791
income       0.3544624  0.6122586
```

Based on the results, I would say that the model is pretty accurate. The p value is very small which means that income does impact food expenditure however I don’t think that the model is a good fit as the r-squared value is only 0.075, meaning income explains just 7.5% of the variation in food spending.

Part c.

As the assignment prescribed I obtained the least squares residuals from the regression in the previous question and plotted them against INCOME. I then created a histogram of the residuals and ran the Jarque-Bera test:






```
> jarque.bera.test(residuals_food_income)

        Jarque Bera Test

data:  residuals_food_income
X-squared = 327.58, df = 2, p-value < 2.2e-16
```

Based on this, I think that the residuals show signs of heteroskedasticity and are not normally distributed for example the scatter plot seems to show that a lot of the residuals are clustered around the bottom left corner of the plot. This suggests that a significant number of low-income households have good expenditures close to or below the model's predicted values. Overall the histogram and the Jarque Bera test indicate that residuals are not normally disturbed. In regression analysis it's more important that the error term is normally distributed to insure valid inference. I think in this case it just indicated that the linear model is not a good fit.

Part d.

As per assignment I created a table which depicts point estimate and the 95% interval estimate of the elasticity of food expenditure with respect to INCOME having the values of 19, 65 and 160.

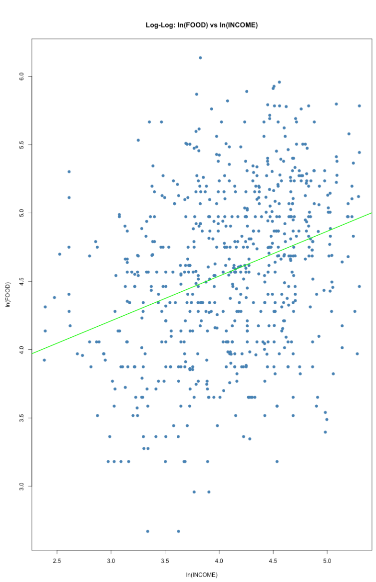
Table: Elasticity Estimates at Different Income Levels

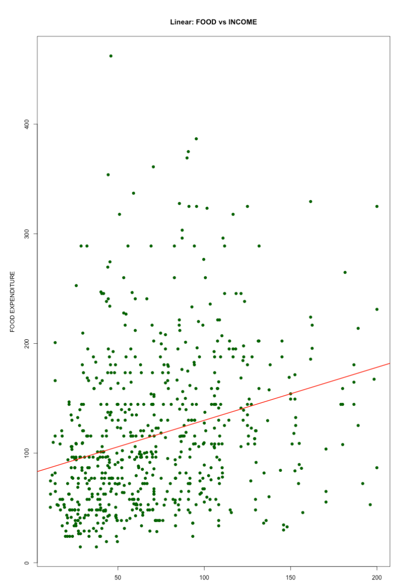
Income	Fitted_FoodExp	Elasticity	Lower_95CI	Upper_95CI
19	90.77	0.1012	0.0662	0.1362
65	113.00	0.2780	0.2008	0.3552
160	158.92	0.4866	0.3906	0.5827

As the table above shows the estimated elasticities are dissimilar and increase with income, ranging from 0.1012 at income 19 to 0.4866 at income 160. Their 95% confidence intervals do not overlap, indicating statistically significant differences. However, based on economic theory (Engel's Law), income elasticity for food should decrease as income rises, since food is a necessity. This suggests the linear model may not fully capture the expected diminishing responsiveness of food spending to income.

Part e.

The log-log model (\ln(\text{FOOD}) \sim \ln(\text{INCOME})) shows a more clearly defined linear relationship compared to the original linear model in levels. In the linear plot, variance in food expenditure increases with income, indicating heteroskedasticity, while the log-log transformation stabilizes the variance and reveals a tighter trend around the fitted line. This suggests that the log-log model fits the data better visually and may be more appropriate for modeling food expenditure with income. To confirm this, comparing the generalized R^2 from the log-log model to the R^2 from the linear model will give a more objective measure of fit.



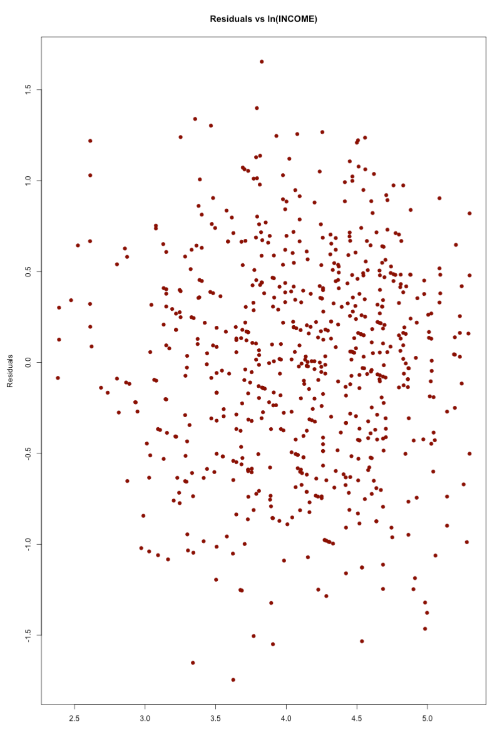


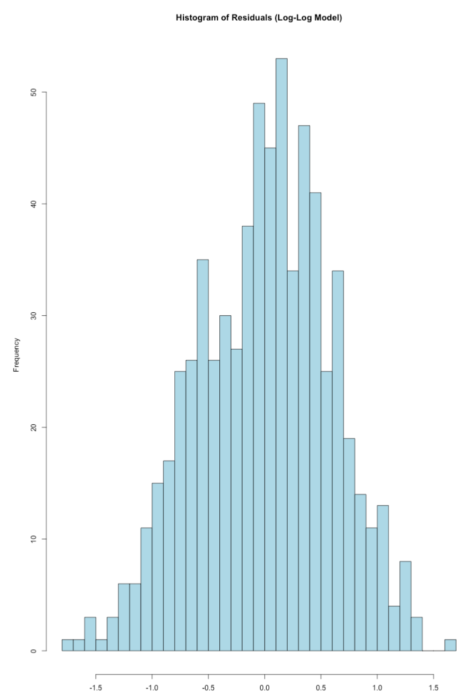
Part f.

The log-log model estimates the elasticity of food expenditure with respect to income at 0.33 (95% CI: 0.25 to 0.40). This elasticity is statistically different from the varying elasticities found in the linear model, as the confidence intervals do not overlap. This provides statistical evidence that the log-log model's elasticity is dissimilar to the income-varying elasticities from the linear model.

```
Table: Elasticity Estimate and 95% CI from Log-Log Model

|          | Estimate| Lower_95CI| Upper_95CI|
|:---------|--------:|----------:|----------:|
|ln_income |   0.3278|     0.2542|     0.4014|
```

Part g.



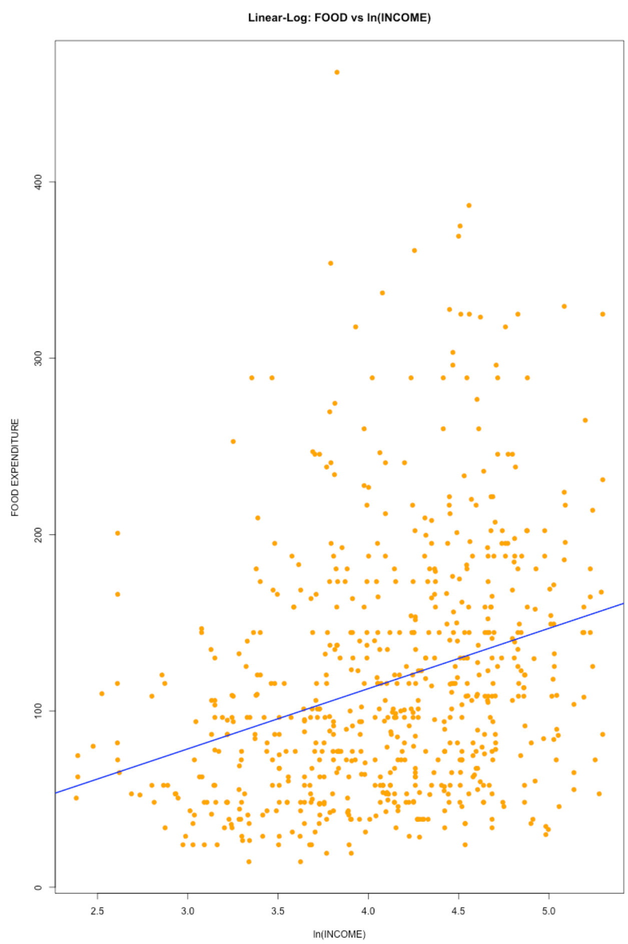



```
        Jarque Bera Test

data:  residuals_loglog
X-squared = 5.5968, df = 2, p-value = 0.06091
```

The residuals from the log-log model show no discernible pattern when plotted against ln(INCOME), supporting the assumption of homoskedasticity. The histogram of residuals is approximately symmetric and bell-shaped. The Jarque–Bera test yields a p-value of 0.06091, so we fail to reject the null hypothesis of normality at the 5% level. Overall, the regression errors in the log-log model appear to be approximately normally distributed.

Part h.

```
Call:
lm(formula = food ~ ln_income, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-116.44  -50.09  -13.76   31.77  355.48

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -24.100     18.029  -1.337    0.182
ln_income     34.193      4.345   7.869 1.44e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 68.14 on 670 degrees of freedom
Multiple R-squared:  0.0846,	Adjusted R-squared:  0.08323
F-statistic: 61.92 on 1 and 670 DF,  p-value: 1.439e-14
```




```
Table: Comparison of R-squared Values Across Models

|Model                          | R_squared|
|:------------------------------|---------:|
|Linear (FOOD ~ INCOME)         | 0.0748598|
|Log-Log (ln(FOOD) ~ ln(INCOME)) | 0.0221569|
|Linear-Log (FOOD ~ ln(INCOME))  | 0.0845988|
```

The linear-log model demonstrates a moderately well-defined relationship between food expenditure and ln(INCOME), with a clear positive trend. Its R^2 = 0.0846 is higher than both the linear model (0.0749) and the log-log model (0.0222), indicating a better fit in terms of explained variance. Visually, however, the linear model may appear more linear, while the log-log model suffers from substantial spread. Overall, the linear-log model fits the data statistically best, though all models explain only a small portion of the variation in food expenditure.

Part i.

The elasticity estimates from the linear-log model are statistically dissimilar to those from both the linear and log-log models. They are much larger in magnitude and are accompanied by wide confidence intervals, some of which even include negative values. This suggests that the linear-log model provides unstable and imprecise elasticity estimates, unlike the more stable and statistically significant estimates from the other models. Therefore, based on both the values and the confidence intervals, the elasticity from the linear-log model is clearly not comparable to the others.

Table: Elasticity Estimates from Linear-Log Model

Income	Fitted_FoodExp	Elasticity	Lower_95CI	Upper_95CI
19	76.58	8.4836	1.1449	15.8223
65	118.64	18.7343	-0.6679	38.1366
160	149.44	36.6103	-7.6622	80.8828