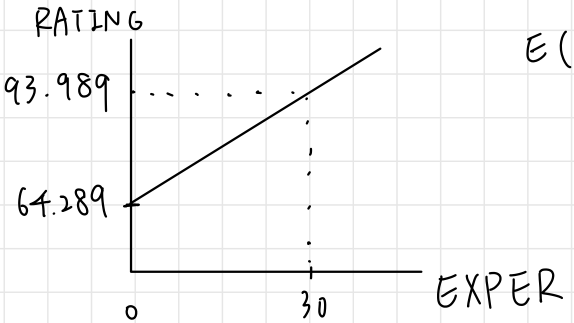
Q4.4.a

For Model 1,

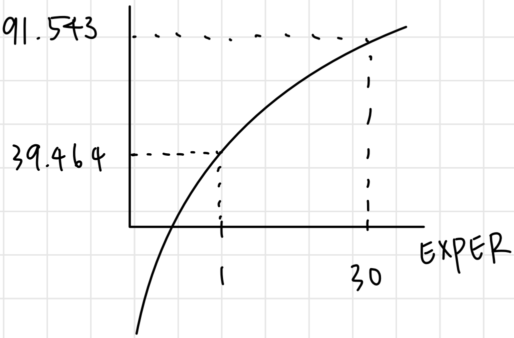


$E(Rating \mid Exper=0) = 64.289$

$E(Rating \mid Exper=30) = 93.989$

Q4.4.b

For Model 2,



$E(Rating \mid Exper=1) = 39.464$

$E(Rating \mid Exper=30) = 91.543$

When $Exper \to 0$, $\widehat{Rating} \to -\infty$, which doesn't have a reasonable economic explanation.

Q4.4.c

M.E.

$$\frac{\partial E(Rating \mid Exper)}{\partial Exper} = 0.99$$

In model one, the marginal effects on rating for any given experience (10y or 20y) are all fixed at 0.99#

Q4.4.d

$$\frac{\partial E(Rating \mid Exper)}{\partial Exper} = 15.312 \times \frac{1}{EXPER}$$

$$\frac{\partial E(Rating \mid Exper=10)}{\partial EXPER} = \frac{15.312}{10} = 1.5312 \ \#$$

$$\frac{\partial E(Rating \mid Exper=20)}{\partial EXPER} = \frac{15.312}{20} = 0.7656 \ \#$$

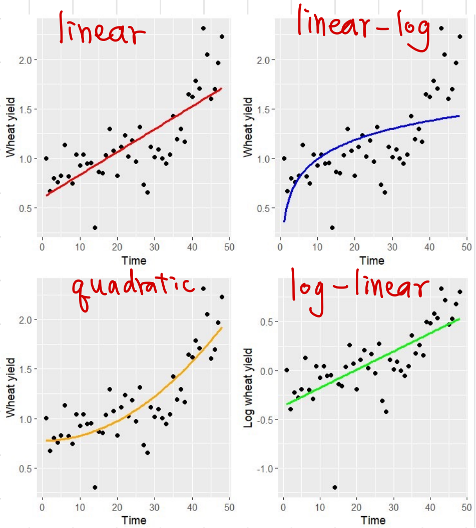
Q4.4.e

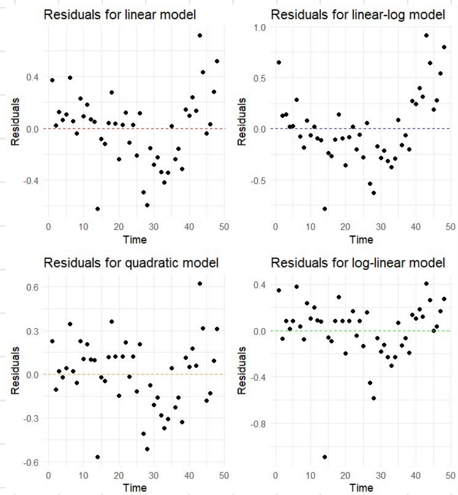
$R_2^2 = 0.6414 > R_1^2 = 0.4858 \rightarrow$ Model 2 fits better

Q4.4.f

Economically, the experience year exhibt a marginal decreasing effect on work quality becaus the learning curve increases in a decreasing way.

Q4.28.a





```
> jarque.bera.test(resid(mod1))

	Jarque Bera Test

data:  resid(mod1)
X-squared = 0.13257, df = 2, p-value =
0.9359
```

linear

```
> jarque.bera.test(resid(mod2))

	Jarque Bera Test

data:  resid(mod2)
X-squared = 2.7629, df = 2, p-value = 0.2512
```

linear-log

```
> jarque.bera.test(resid(mod3))

	Jarque Bera Test

data:  resid(mod3)
X-squared = 0.32406, df = 2, p-value =
0.8504
```

quadratic

```
> jarque.bera.test(resid(mod4))

	Jarque Bera Test

data:  resid(mod4)
X-squared = 83.874, df = 2, p-value <
2.2e-16
```

log-linear

rej. normal error

```
> 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
```

According to JB test, with 5% confidence, the errors of linear, linear-log, and quadratic models are normally distributed.

The R^2 of the quadratic model is the highest and it's residuals pass the JB test. Thus, I think the quadratic model is the best choice.

Q4.4.b In the quadratic model,

```
> coef(mod3)[2]
   I(time^2)
0.0004986181
```

$$\frac{\partial \widehat{Yield}}{\partial Time} = 2\gamma_1 Time$$

, meaning the marginal effect that time did on yield will be affected by γ_1.

$\hat{\gamma}_1 = 0.0005$, meaning its a convex function.

Q4.4.c The following observations are unusual under various criteria.

	Outliers	High_Leverage	Influential_DFBETAS..Intercept.	Influential_DFBETAS.I.time.2.	Influential_DFFITS
14	TRUE	FALSE	FALSE	FALSE	TRUE
28	TRUE	FALSE	FALSE	FALSE	FALSE
43	TRUE	FALSE	FALSE	FALSE	TRUE
45	FALSE	TRUE	FALSE	FALSE	FALSE
46	FALSE	TRUE	FALSE	FALSE	FALSE
47	FALSE	TRUE	FALSE	FALSE	FALSE
48	FALSE	TRUE	FALSE	FALSE	TRUE

Q4.4.d

Below is the prediction interval of Time = 48 (1997) trained by observations up to 1996

```
     fit      lwr      upr
1.881111 1.372403 2.389819
```

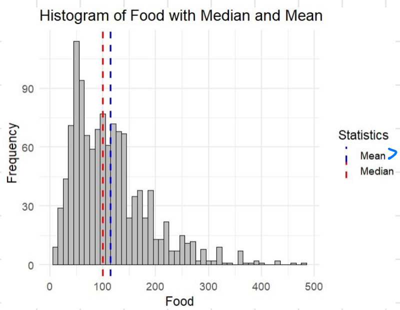
	northampton
41	1.7875
42	1.7104
43	2.3161
44	2.0534
45	1.6040
46	1.6980
47	1.9691
48	2.2318

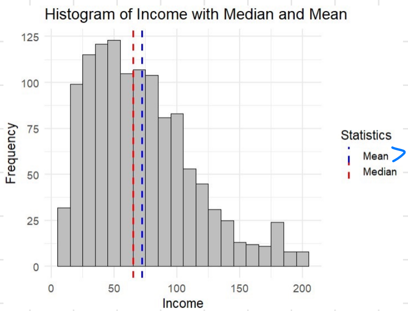
The real observation of yield in 1997 is 2.2318 $\in$ Prediction interval $[1.3724, 2.39]$ #

Q4.29.a

```
summary(data$food)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 9.63   57.78   99.80  114.44  145.00  476.67
summary(data$income)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
10.00   40.00   65.29   72.14   96.79  200.00
```

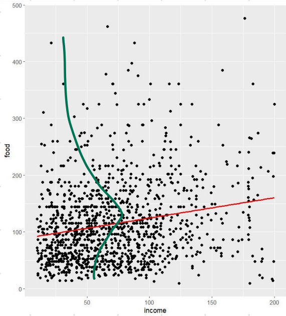
```
        Jarque Bera Test

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

        Jarque Bera Test

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





Both data are rejected to follow normal dist. and skewed to the right.

Q4.29.b

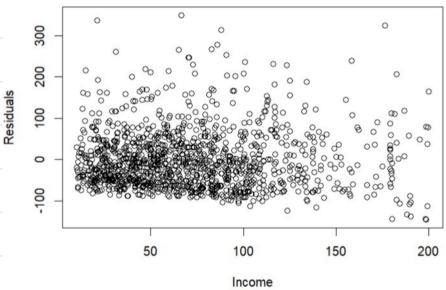


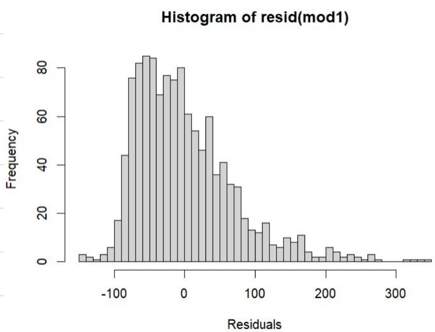
		2.5 %	97.5 %
(Intercept)	β_1	80.5064570	96.626543
income	β_2	0.2619215	0.455452

The residuals seems to be skewed. However, the sample size is large enough to make approximation of estimator by CLT. Thus, the interval estimate is still precise.

* $n=1200$ here !! $\hookrightarrow t(n) \xrightarrow{a} Z$

Q4.29.c






```
> jarque.bera.test(resid(mod1))

        Jarque Bera Test

data:  resid(mod1)
X-squared = 624.19, df = 2, p-value <
2.2e-16
```

The residuals skewed to the right and the JB test result reject the null that the error is normally distributed.

It is more important to make sure the error term follows normality especially in small sample inference.

When the error is non-normal, we cannot find the actual distribution of $t = \frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$, and thus we can't construct interval estimate and statistical inference like t-test and F-test. to testify β and σ^2.

Q4.29.d

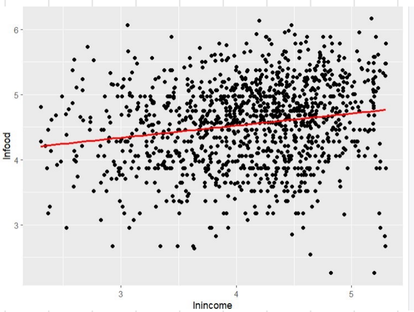
	INCOME	Fitted_Elasticity	Lower_Bound	Upper_Bound
1	19	0.07145038	0.05219423	0.09070654
2	65	0.20838756	0.15222630	0.26454882
3	160	0.39319883	0.28723022	0.49916745

$$\frac{\partial \widehat{Food}/\widehat{Food}}{\partial income/income} = \hat{\beta}_2 \times \frac{income}{\widehat{Food}}$$

$$SE(elas) = SE(\hat{\beta}_2) \times \frac{income}{\widehat{Food}}$$

The bonds do not overlap and the fitted elasticity increase as income increases, meaning that food might be some kind of luxuary. Families with higher income tend to increase more spending on food than those with lower income.

Q4.29.e



Generalized $r^2 = Cor(y_i, \hat{y}_i)^2$

The linear model have higher goodness of fit in terms of generalized R^2 than the log-log model.

```
> rg1
[1] 0.0422812
> rg2
[1] 0.03965161
```

linear

log-log

Q4.29.f

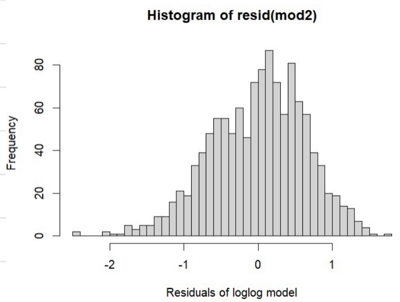
```
> ci2
                2.5 %    97.5 %
(Intercept) 3.5428135 4.0150507
lnincome    0.1293432 0.2432675
```

$$\ln(\widehat{Food}) = \gamma_1 + \gamma_2 \ln(Income)$$

$$\hat{\varepsilon} = \frac{\partial \ln Food}{\partial \ln Income} = \hat{\gamma}_2$$

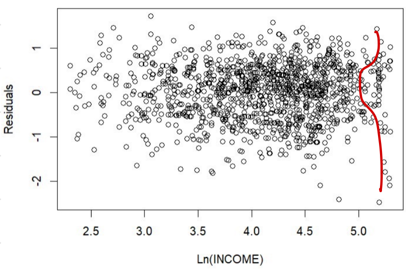
95% CI for $\hat{\varepsilon} = [0.129, 0.243]$ which is fixed for the log-log model.

Q4.29 g.



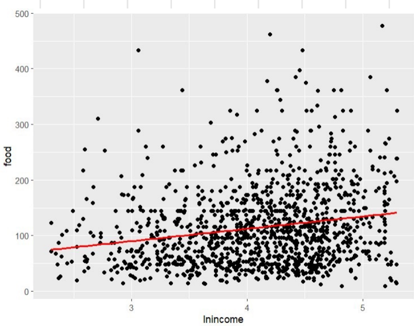

```
Jarque Bera Test

data:  resid(mod2)
X-squared = 25.85, df = 2, p-value =
2.436e-06
```

According to JB test, it rejected the error of log-log model is normally distributed



The residuals skew to the right.

Q4.29. h For a linear-log model, it has a lower R^2 than linear model but higher R^2 than log-log model.




```
> summary(mod1)$r.squared
[1] 0.0422812
> summary(mod2)$r.squared
[1] 0.03322915
> summary(mod3)$r.squared
[1] 0.03799984
```

linear

log-log

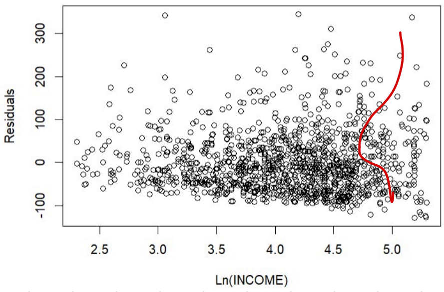
linear-log

Q4.29. i

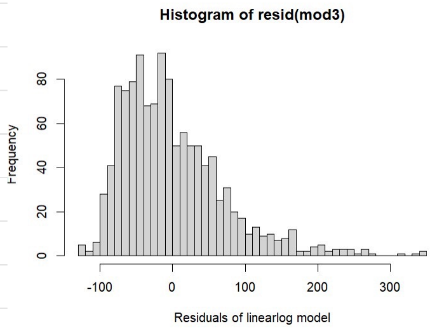
	LN_INCOME	Fitted_Elasticity	Lower_Bound	Upper_Bound
1	2.944439	0.2495828	0.1784728	0.3206929
2	4.174387	0.1909624	0.1365542	0.2453705
3	5.075174	0.1629349	0.1165122	0.2093576

The elastisity decreases as ln Income ↓

Q4.29. j




```
        Jarque Bera Test

data:  resid(mod3)
X-squared = 628.07, df = 2, p-value < 2.2e-16
```



The residuals skew to the left and the JB test reject the error of the linear-log model is normally distributed

Q4.29. k

I would prefer linear model due to its highest R^2. One might argue that the error of the log-log model is most likely to be drawn from normality compared to the others. But having 1200 observations allows us to use approximation on the sampling distribution to make furthur inference. Therefore, the non-normal error is relatively subtle than the goodness of fit here.