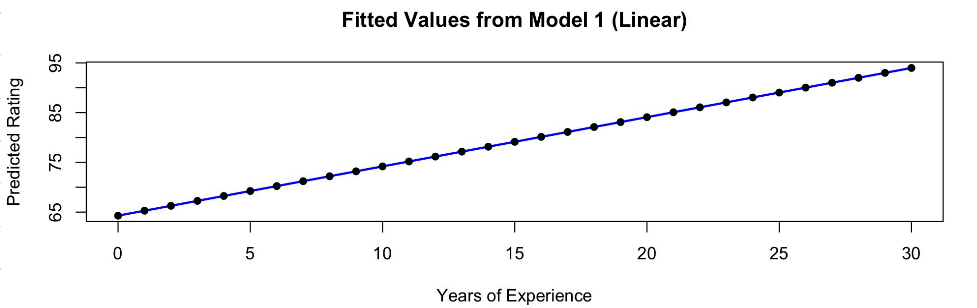
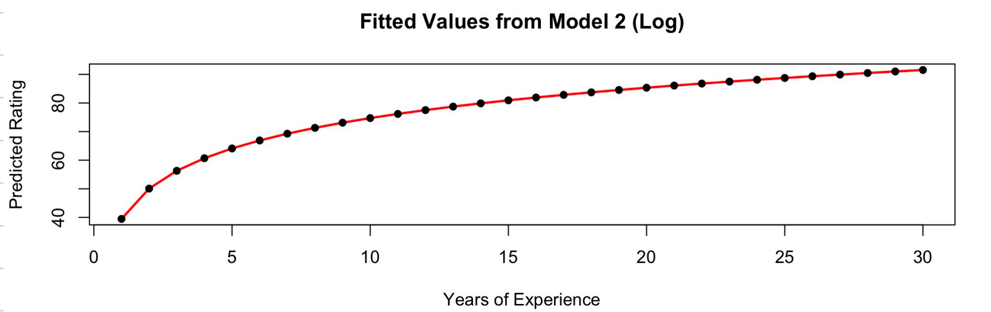
4.04

a.



b.



c.

Model 1: $\widehat{RATING} = 64.289 + 0.99\,EXPER$

marginal effect: $\dfrac{\partial\, RATING}{\partial\, EXPER} = 0.99$ # for 10 years and 20 years d、

Model 2: $\widehat{RATING} = 39.464 + 15.312\ln(EXPER)$

marginal effect: $\frac{\partial \widehat{RATING}}{\partial \ln(EXPER)} = \frac{15.312}{EXPER}$

(i) 10 years:

$\frac{15.312}{10} = 1.5312$ #

(ii) 20 years

$\frac{15.312}{20} = 0.7656$ #

e、

∵ Model 2's $R^2 = 0.6414$ > Model 1's $R^2 = 0.4858$

∴ Model 2 is a better Model.

f、

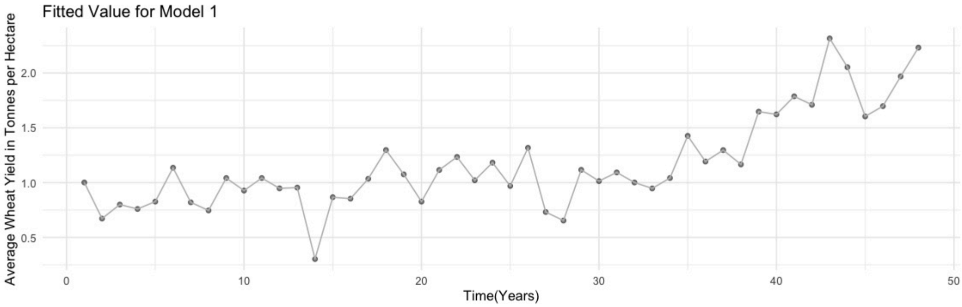
Model 2 應該是比較合理的

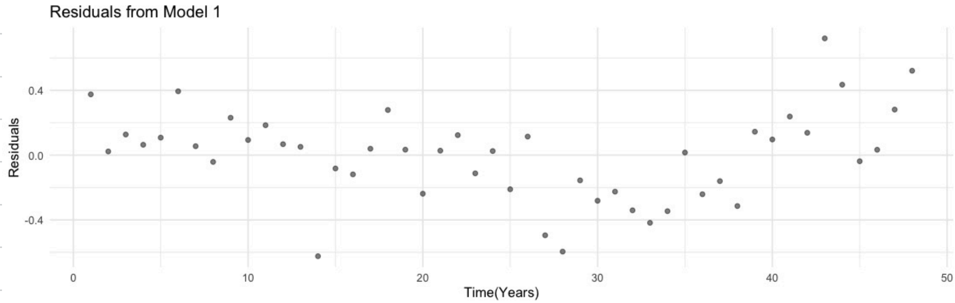
∵ 隨著經驗逐漸增加，表現會愈好，但當經驗豐富的人 (maybe 20 years ↑) 增加1年經驗對於表現影響不大 (邊際遞減)

4.28

a、

Model 1: $YIELD_t = \beta_0 + \beta_1 TIME + e_t$

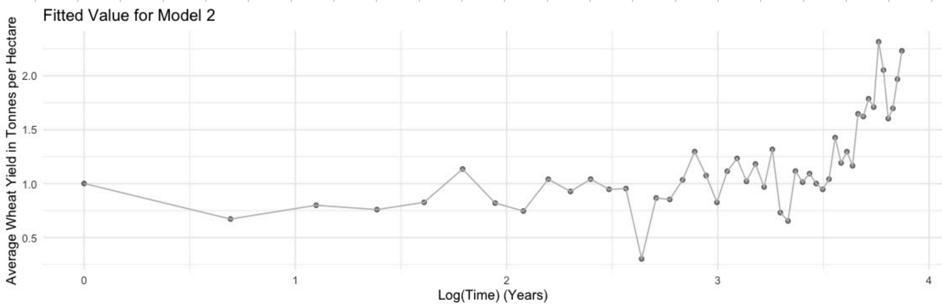


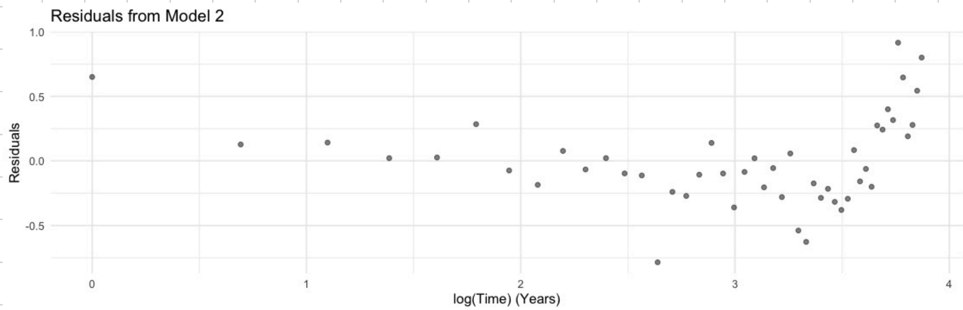


Shapiro-Wilk nomality test: $W = 0.9824$ P-value $= 0.6792$

$R^2 = 0.5778$

Model 2: $YIELD_t = \alpha_0 + \alpha_1 \ln(TIME) + e_t$

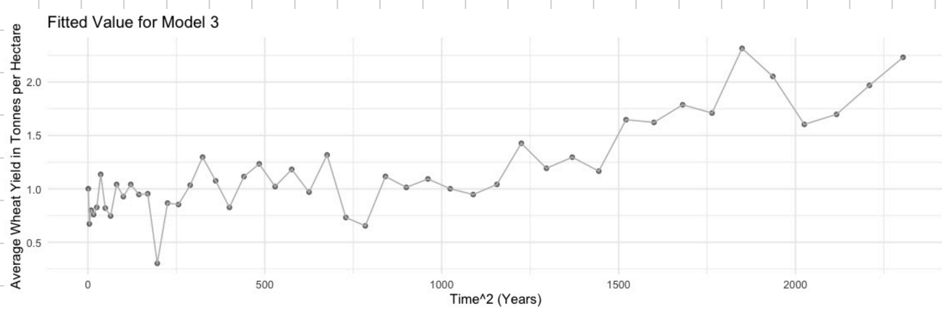


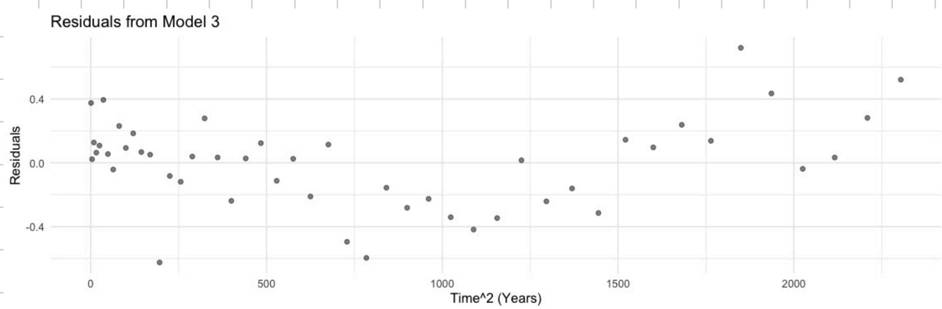


Shapiro-Wilk nomality test: $W = 0.9666$ $p\text{-value} = 0.1856$

$R^2 = 0.3386$

Model 3: $YIELD_t = r_0 + r_1 TIME^2 + e_t$

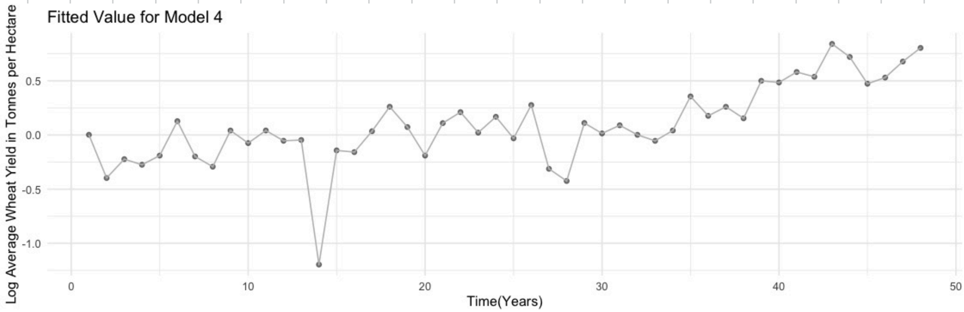


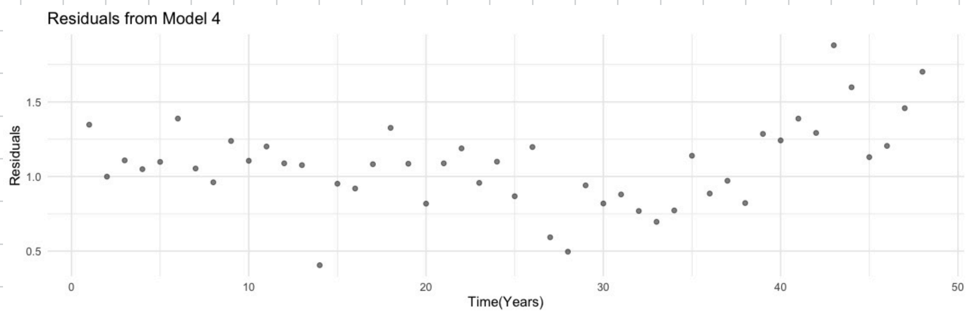


Shapiro-Wilk nomality test: $W = 0.9824$ $p\text{-value} = 0.6792$

$R^2 = 0.5778$

Model 4: $\ln(YIELD_t) = \beta_0 + \beta_1 TIME + e_t$





Shapiro-Wilk normality test: $W=0.8689$

p-value $= 7.205e-05$ $R^2=0.5074$

Choose Model 3 $\because$ R^2 最大

b.

Model 3: $YIELD_t = \gamma_0 + \gamma_1 TIME^2 + e_t$

$$\frac{\partial YIELD_t}{\partial TIME} = 2\gamma_1 TIME$$

當TIME增加1單位，以平均來說，YIELD會增加 $2\gamma_1 TIME$ 單位

c-

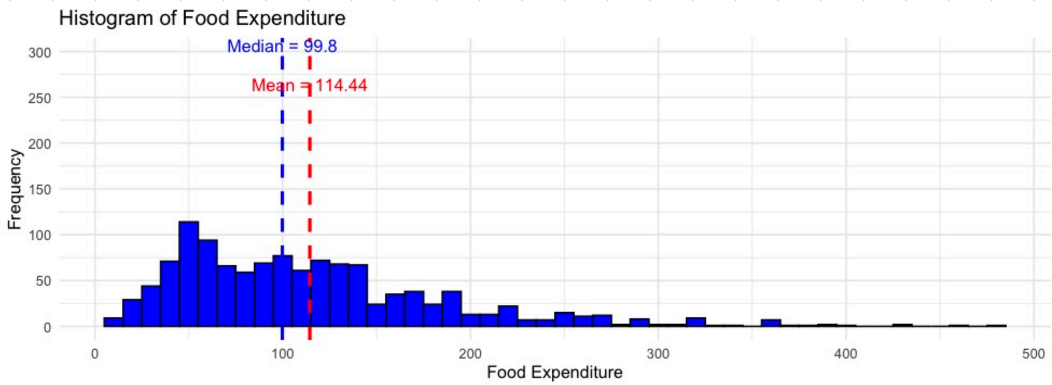
```
Unusual observations based on Studentized Residuals (>| 2 |):
> print(unusual_resid)
14 28 43 48
14 28 43 48
> cat("\nUnusual observations based on Leverage (>", round(threshold_leverage, 3),
"):\n")

Unusual observations based on Leverage (> 0.083 ):
> print(unusual_leverage)
named integer(0)
> cat("\nUnusual observations based on DFBETAS (>|", round(threshold_dfbetas, 3),
"| for any coefficient):\n")

Unusual observations based on DFBETAS (>| 0.289 | for any coefficient):
> print(unusual_dfbetas)
 1  6 14 43 44 48
 1  6 14 43 44 48
> cat("\nUnusual observations based on DFFITS (>|", round(threshold_dffits, 3),
"|):\n")

Unusual observations based on DFFITS (>| 0.408 |):
> print(unusual_dffits)
 1 14 43 44 48
 1 14 43 44 48
```

d-

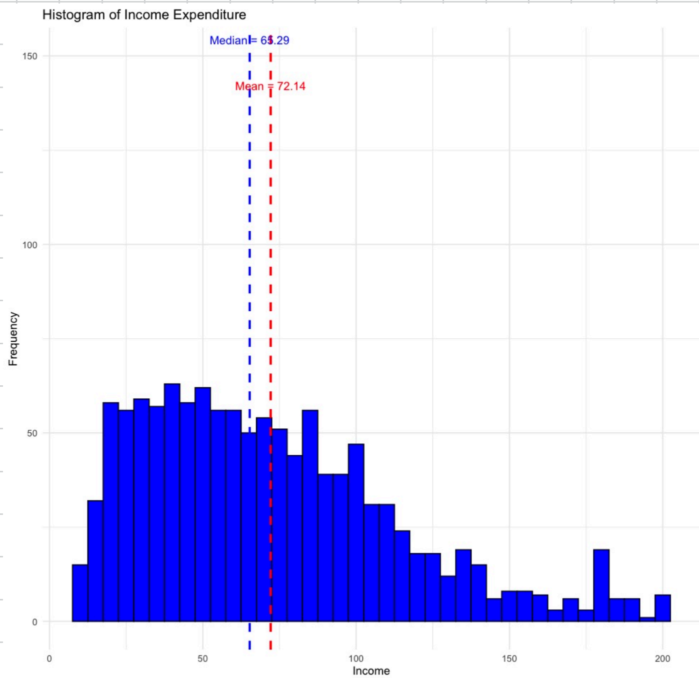
```
> data_train <- subset.data.frame(wa_wheat, select = time < 48,)
> # Fit your chosen model; here we use a log-linear model as an example.
> model5 <- lm(northampton ~ time^2, data = data_train)
> summary(model5)

Call:
lm(formula = northampton ~ time^2, data = data_train)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.62394 -0.17302  0.03342  0.12996  0.72050 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.603245   0.081858   7.369 2.55e-09 ***
time        0.023078   0.002908   7.935 3.69e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2791 on 46 degrees of freedom
Multiple R-squared:  0.5778,	Adjusted R-squared:  0.5687 
F-statistic: 62.96 on 1 and 46 DF,  p-value: 3.689e-10

> # Create new data for prediction for 1997, assuming time = 48 corresponds to 199
7
> new_data <- data.frame(time = 48)
> # Construct a 95% prediction interval
> pred <- predict(model5, newdata = new_data, interval = "prediction", level = 0.9
5)
> # If the model is in log-scale, transform predictions back:
> pred_exp <- exp(pred)
> print(pred_exp)
       fit      lwr      upr
1 5.534412 3.085882 9.925755
> # Get the true yield for 1997 (time = 48)
> true_value <- wa_wheat$northampton[wa_wheat$time == 48]
> cat("True yield for 1997:", true_value, "\n")
True yield for 1997: 2.2318 
> cat("95% prediction interval for 1997 (exponentiated): [", pred_exp[,"lwr"],
",", pred_exp[,"upr"], "]\n")
95% prediction interval for 1997 (exponentiated): [ 3.085882 , 9.925755 ]
> 
> if (true_value >= pred_exp[,"lwr"] && true_value <= pred_exp[,"upr"]) {
+   cat("The true value is contained in the prediction interval.\n")
+ } else {
+   cat("The true value is NOT contained in the prediction interval.\n")
+ }
The true value is NOT contained in the prediction interval.
```

4:29

a-



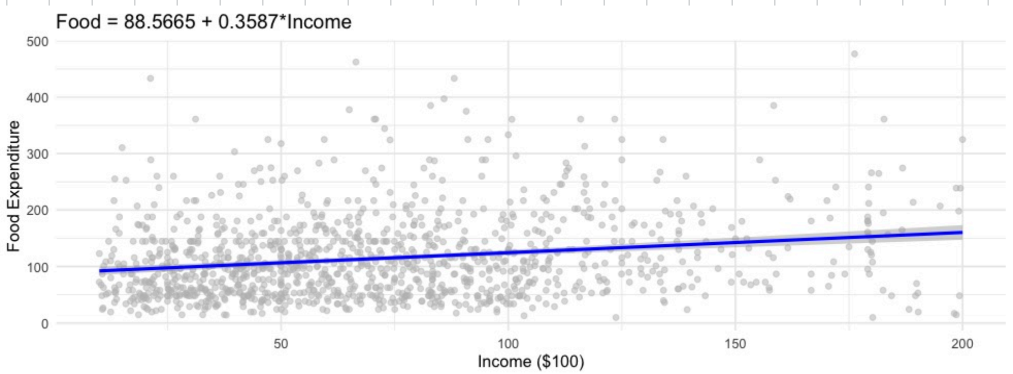
mean_food	median_food	maximum_food	minimum_food	sd_food
114.4431	99.8	476.67	9.63	72.6575



```
| mean_income| median_income| maximum_income| minimum_income| sd_income|
|-----------:|-------------:|--------------:|--------------:|---------:|
|    72.14264|         65.29|            200|             10|  41.65228|
```

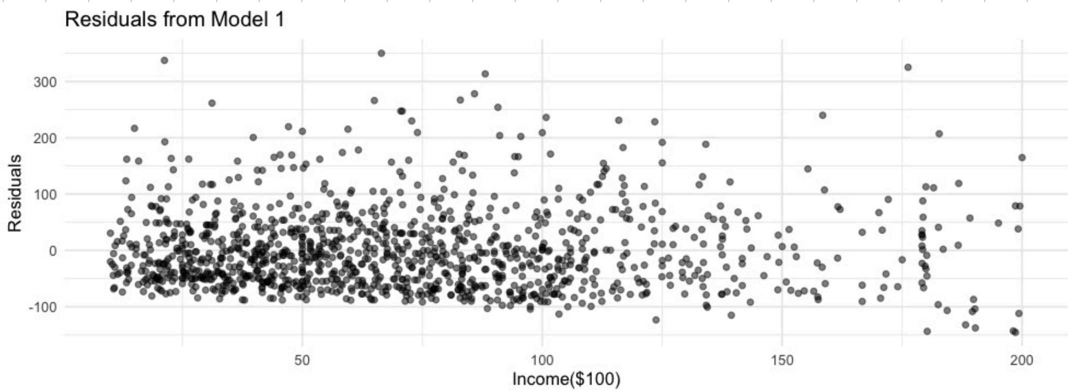
```
	Jarque Bera Test

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

> # Jarque-Bera test for INCOME
> jb_income <- jarque.bera.test(cex5_small$income)
> print(jb_income)

	Jarque Bera Test

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

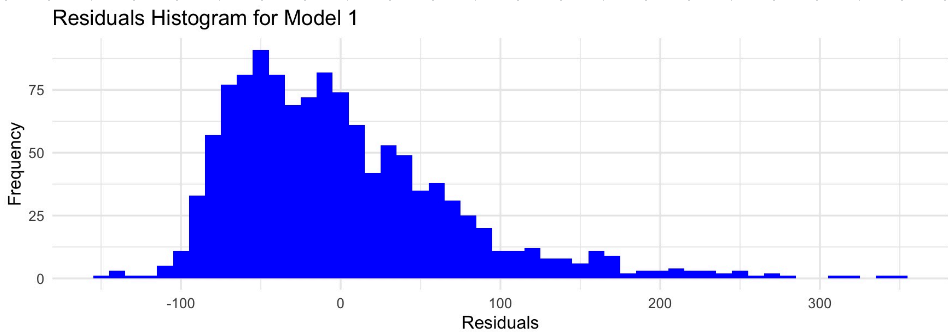
b-



95% C.I for β_2: [0.2619, 0.4555]

C.






```
        Jarque Bera Test

data:  residuals1
X-squared = 624.19, df = 2, p-value < 2.2e-16
```

d.

Model: $FOOD = 88.5665 + 0.3587 \times Income$

95% C.I: $[0.2619, 0.4555]$

elasticity: $\dfrac{\partial FOOD/FOOD}{\partial Income/Income} = \dfrac{\partial FOOD}{\partial Income} \times \dfrac{Income}{FOOD} = 0.3587 \times \dfrac{Income}{FOOD}$

當 income=19, FOOD=95.3818

$\varepsilon = 0.3587 \times \dfrac{19}{95.3818} = 0.0715$

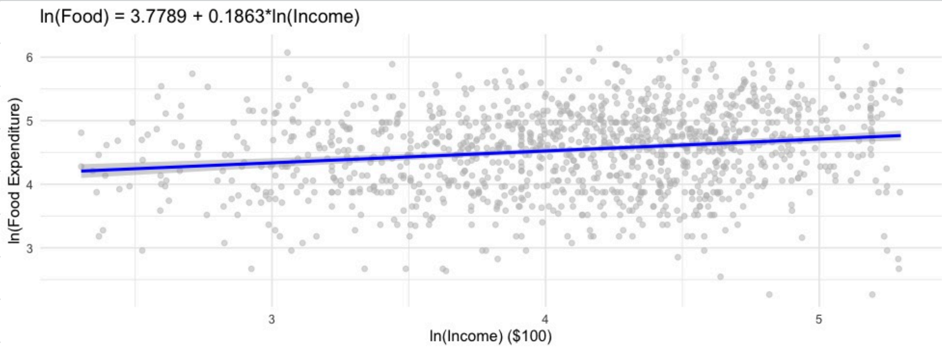
當 Income=65, FOOD=111.882

$\varepsilon = 0.3587 \times \dfrac{65}{111.882} = 0.2084$

當 Income=160, FOOD=145.9585

$\varepsilon = 0.3587 \times \dfrac{160}{145.9585} = 0.3932$

e.




```
Call:
lm(formula = food ~ income, data = cex5_small)

Residuals:
    Min      1Q  Median      3Q     Max 
-145.37  -51.48  -13.52   35.50  349.81 

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
```

```
Call:
lm(formula = log(food) ~ log(income), data = cex5_small)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.48175 -0.45497  0.06151  0.46063  1.72315 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.77893    0.12035  31.400   <2e-16 ***
log(income)  0.18631    0.02903   6.417    2e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 0.6418 on 1198 degrees of freedom
Multiple R-squared:  0.03323,	Adjusted R-squared:  0.03242 
F-statistic: 41.18 on 1 and 1198 DF,  p-value: 1.999e-10
```

linear model is better because the R^2 is larger.

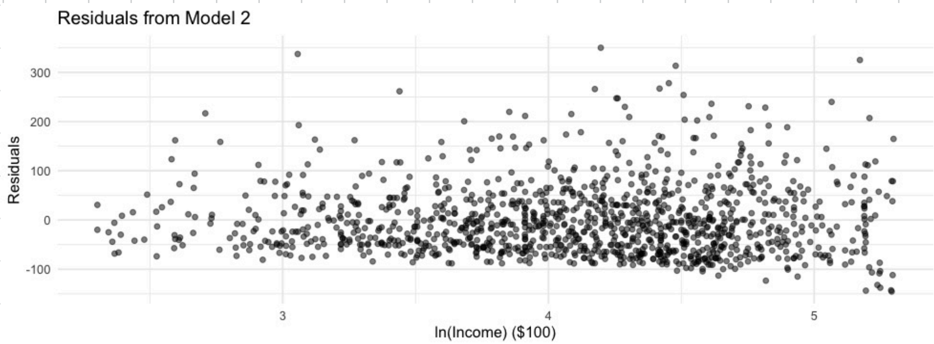
f -

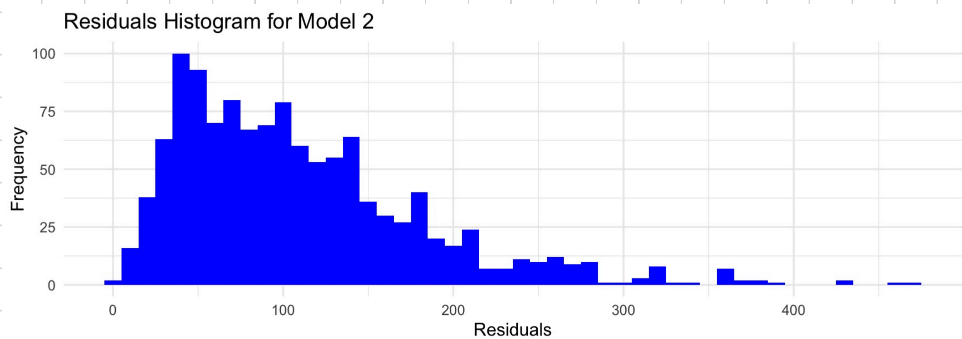
$$\ln(\text{Food}) = 3.7789 + 0.1863 \ln(\text{income})$$

elasticity = 0.1863

95% C.I: [0.1293, 0.2433]

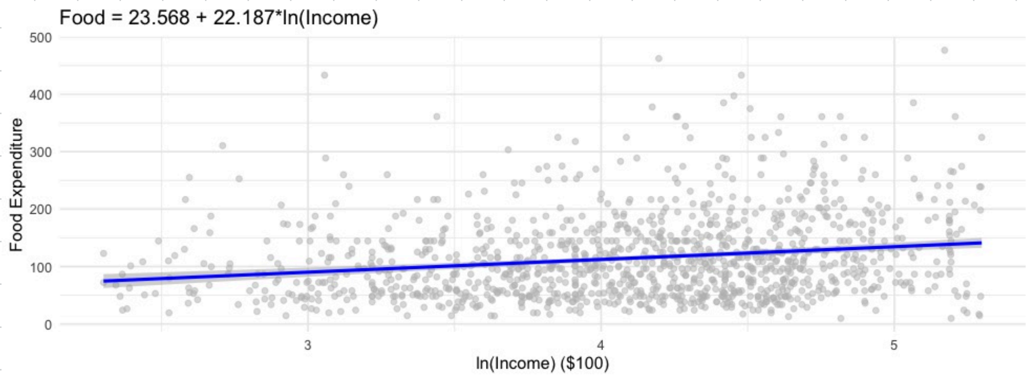
g -






```
        Jarque Bera Test

data:  residuals2
X-squared = 649.18, df = 2, p-value < 2.2e-16
```

h.



i

Model: $Food = 23.568 + 22.187\ \ln(Income)$

95% C.I $[15.8595, 28.5153]$

elasticity $\varepsilon = \dfrac{\partial Food / Food}{\partial \ln(Income) / Income} = \dfrac{22.187}{y}$

當 income = 19 時, Food = 51.9397

$\varepsilon = \dfrac{22.187}{51.9397} = 0.4272$

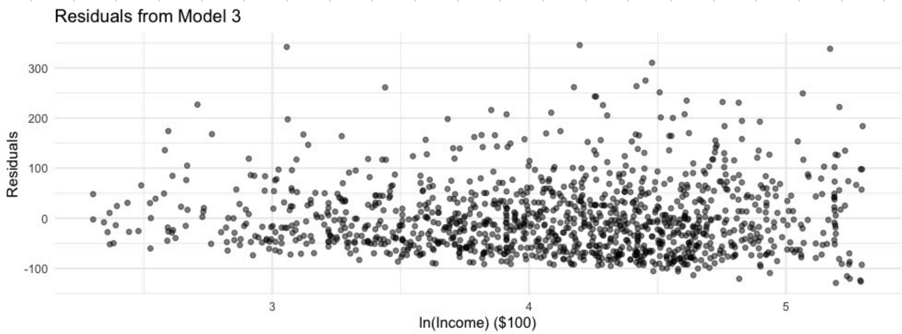
當 Income = 65 時, Food = 63.7911

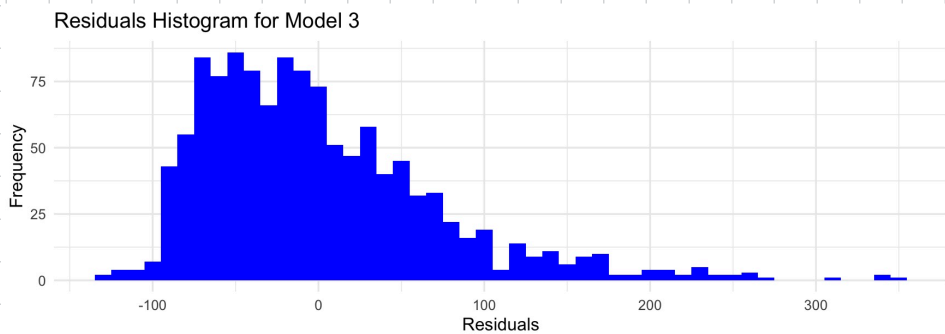
$\varepsilon = \dfrac{22.187}{63.7911} = 0.3478$

當 Income = 160 時, Food = 72.4708

$\varepsilon = \dfrac{22.187}{72.4708} = 0.3062$

j.






```
        Jarque Bera Test

data:  residuals3
X-squared = 628.07, df = 2, p-value < 2.2e-16
```

k.

prefer log-log，由殘差直方圖與散佈圖可以看出較像鐘形曲線，且平均數接近0