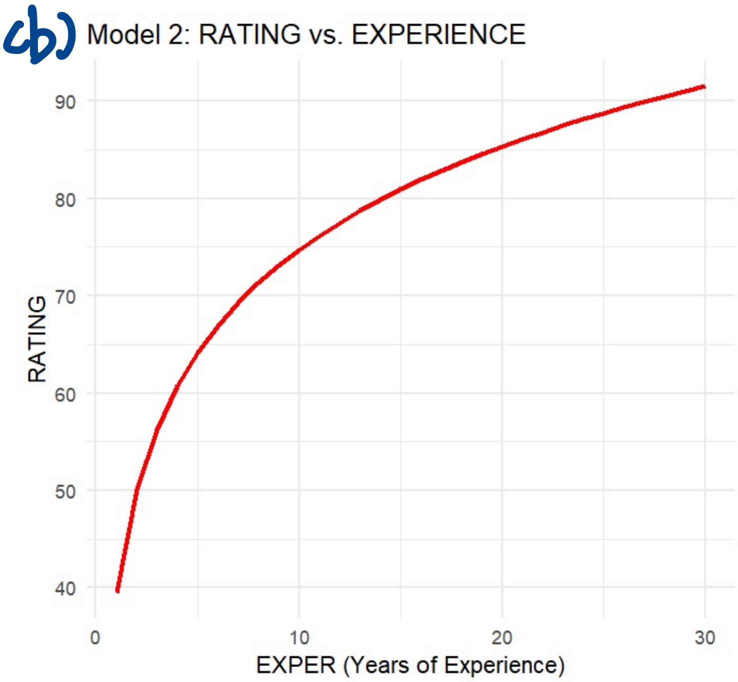
4.4

因 ln0 不存在
⇒ artists with no experience 不在此 model

(a)

(b)



(c) model 1 是線性的 ⇒ margin effect 會一樣 = 斜率 = 0.990

(d) exper = 10 ⇒ margin effect = $\frac{15.312}{10}$ = 1.5312

exper = 20 ⇒ margin effect = $\frac{15.312}{20}$ = 0.7656

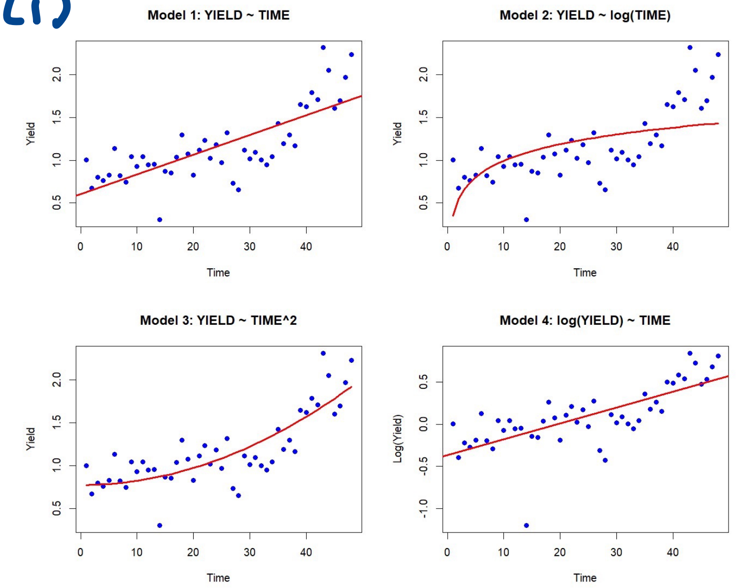
(e) 若用 R-square 比較兩 model, model 2 > model 1
⇒ model 2 fits the data better

(f) Model 2 is more resonable, 比較符合現實情形.
也反應了 exper 對於 performance 的 margin effect 遞減.

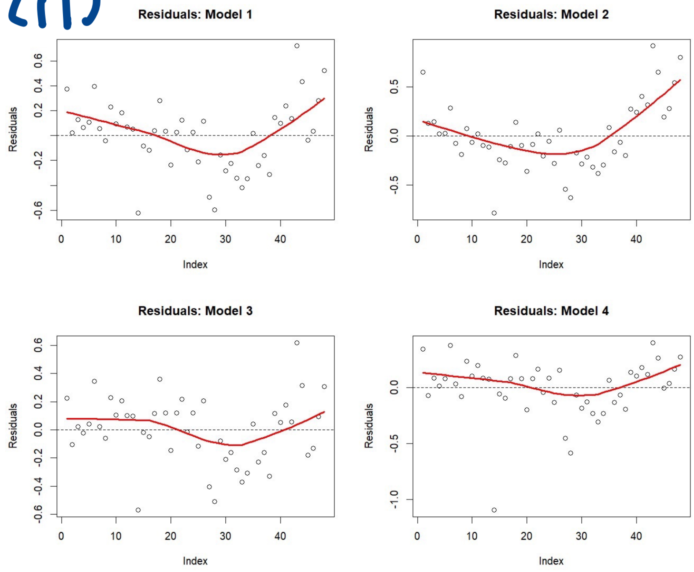
4.28

(a)

(i)



(ii)



(iii) Shapiro-Wilk normality test

```
data:  model1$residuals
W = 0.98236, p-value = 0.6792
data:  model2$residuals
W = 0.96657, p-value = 0.1856
data:  model3$residuals
W = 0.98589, p-value = 0.8266
data:  model4$residuals
W = 0.86894, p-value = 7.205e-05
```

0.6792 > 0.05

0.1856 > 0.05

0.8266 > 0.05

7.205e-05 < 0.05

model 3 is preferable, because it has the highest R-squared and P-value of error normality test.

(iv) R-squared:

Linear	Log(TIME)	TIME^2	Log(YIELD)
0.5778369	0.3385733	0.6890101	0.5073566

(b)

```
Call:
lm(formula = northampton ~ I(time^2), data = wa_wheat)

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
```

$\widehat{yield} = 0.7737 + 0.0004986\ time^2$

4.986e-04 > 0 → It means yield will grow accelerately over time.

10.10 extremely significant ⇒ time has strong impact on yield

(c)

	time	yield
1	1	1.0014
2	2	0.6721
14	14	0.3024
28	28	0.6539
43	43	2.3161
47	47	1.9691
48	48	2.2318

(d)

	fit	lwr	upr
1	1.8811	1.3724	2.3898

95% CI = [1.3724, 2.3898]

Actual northampton yield in 1997: 2.2318 in the 95% CI

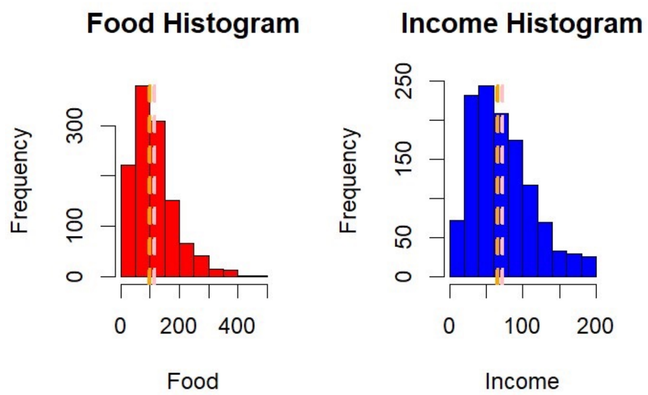
The actual yield is within the 95% prediction interval.

4.29

(a)

```
> cat("Income Summary:\n")
Income Summary:
> cat("Mean: ", income_stats["mean"], "\n")
Mean:  72.14264
> cat("Median: ", income_stats["median"], "\n")
Median:  65.29
> cat("Min: ", income_stats["min"], "\n")
Min:  10
> cat("Max: ", income_stats["max"], "\n")
Max:  200
> cat("Standard Deviation: ", income_stats["sd"], "\n\n")
Standard Deviation:  41.65228
```

```
> cat("Food Summary:\n")
Food Summary:
> cat("Mean: ", food_stats["mean"], "\n")
Mean:  114.4431
> cat("Median: ", food_stats["median"], "\n")
Median:  99.8
> cat("Min: ", food_stats["min"], "\n")
Min:  9.63
> cat("Max: ", food_stats["max"], "\n")
Max:  476.67
> cat("Standard Deviation: ", food_stats["sd"], "\n\n")
Standard Deviation:  72.6575
```



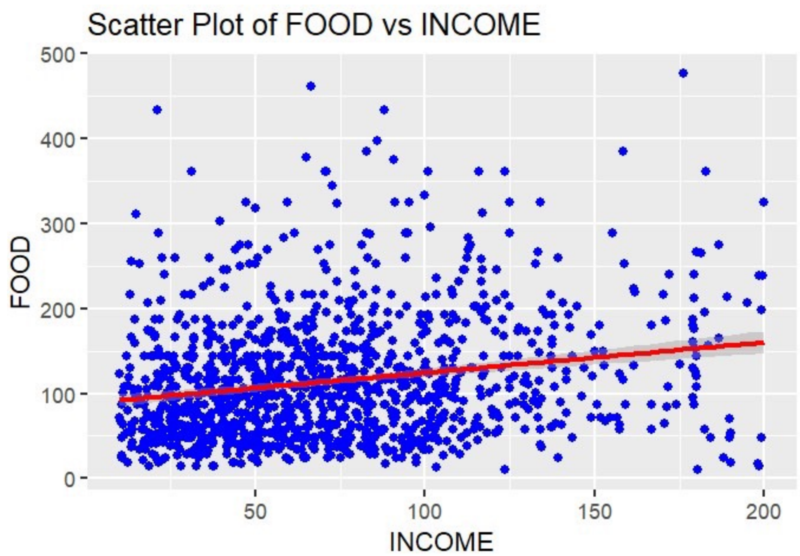
mean > median ⇒ Both aren't bell-shaped.

```
FOOD 的 Jarque-Bera 檢定統計量: 645.6099

INCOME 的 Jarque-Bera 檢定統計量: 147.6768

> cat("INCOME 的 p-value:", p_value_income, "\n")
INCOME 的 p-value: 0
> p_value_food <- 1 - pchisq(JB_food, df=2)
> cat("FOOD 的 p-value:", p_value_food, "\n")
FOOD 的 p-value: 0
```

< 0.05 (INCOME), < 0.05 (FOOD)

⇒ reject H_0, residuals of income and food don't follow a normal distribution.

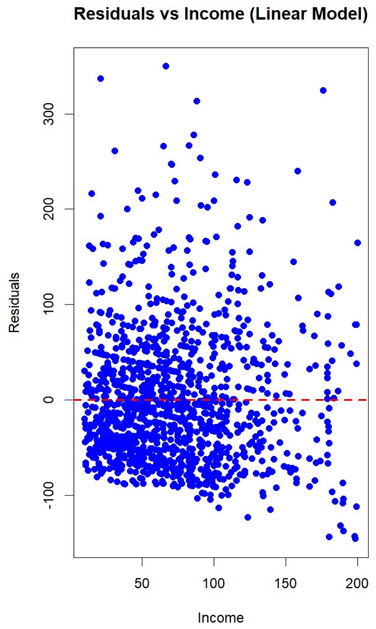
(b)

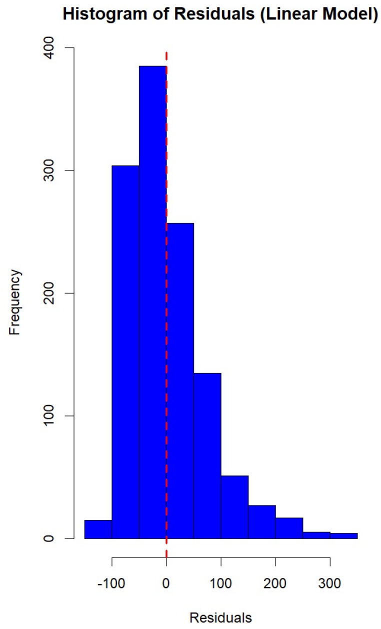



```
                  2.5 %      97.5 %
(Intercept) 80.5064570 96.626543
income       0.2619215  0.455452
```

```
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
```

(c)





Jarque Bera Test :

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

(d)

increasing with income

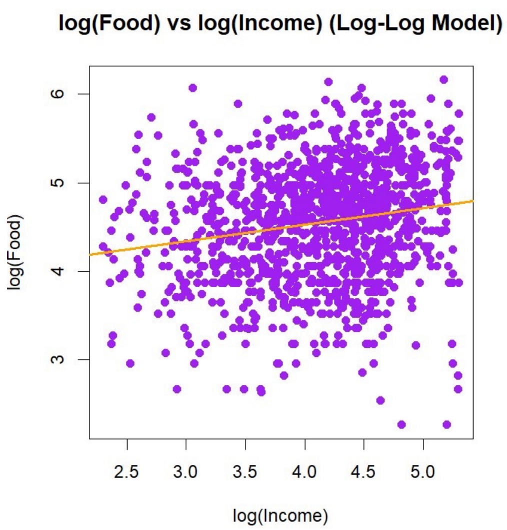
```
  income food_predicted elasticity lower_95_ci upper_95_ci
1     19          95.38     0.0715      0.0522      0.0907
2     65         111.88     0.2084      0.1522      0.2646
3    160         145.96     0.3932      0.2871      0.4993
```

increasing with income

⇒ isn't consistent with standard economic predictions.

CI don't overlap ⇒ there is a significant difference of food expenditure from different income levels.

(e)




```
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 ' ' 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

R-squared of Linear Model:  0.0422812
> cat("R-squared of Log-Log Model: ", r2_loglog, "\n")
R-squared of Log-Log Model:  0.03322915
```

$\ln(FOOD) = 3.77893 + 0.18631 \ln(INCOME)$

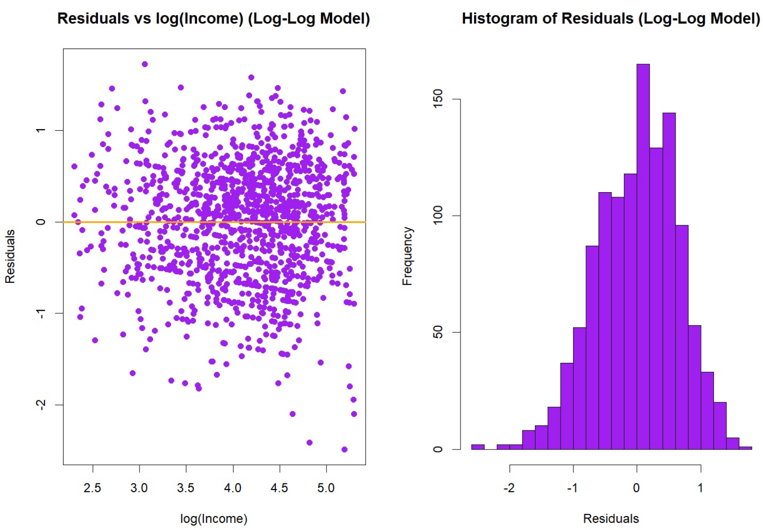
(f)

```
               2.5 %    97.5 %
(Intercept) 3.5428135 4.0150507
log_income  0.1293432 0.2432675
```

The elaticity in log-log model is fixed not increasing with income

⇒ dissimilar with part (d)

(g)



X-squared = 25.85, df = 2

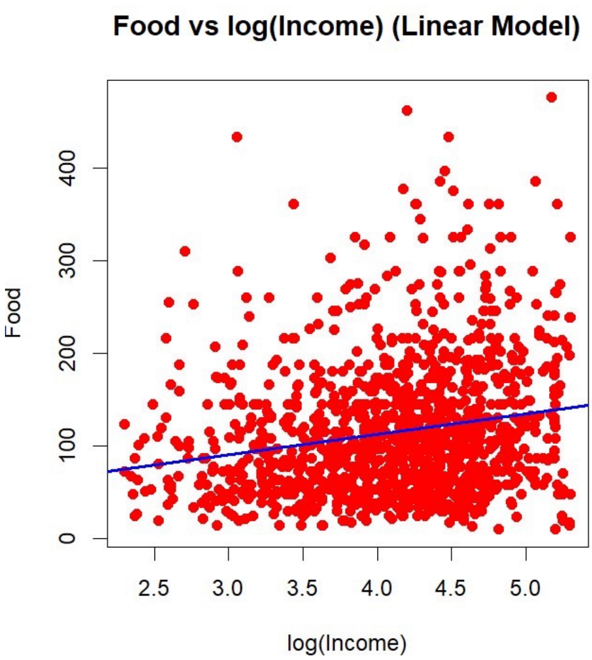
p-value = 2.436e-06

25.85 > 5.99

⇒ reject H_0

(log-log regression errors are normal)

(h)




```
R-squared of Linear Model with log(Income):  0.03799984
> cat("R-squared of Log-Log Model: ", r2_loglog, "\n")
R-squared of Log-Log Model:  0.03322915
```

linear-log is larger than log-log

(i)

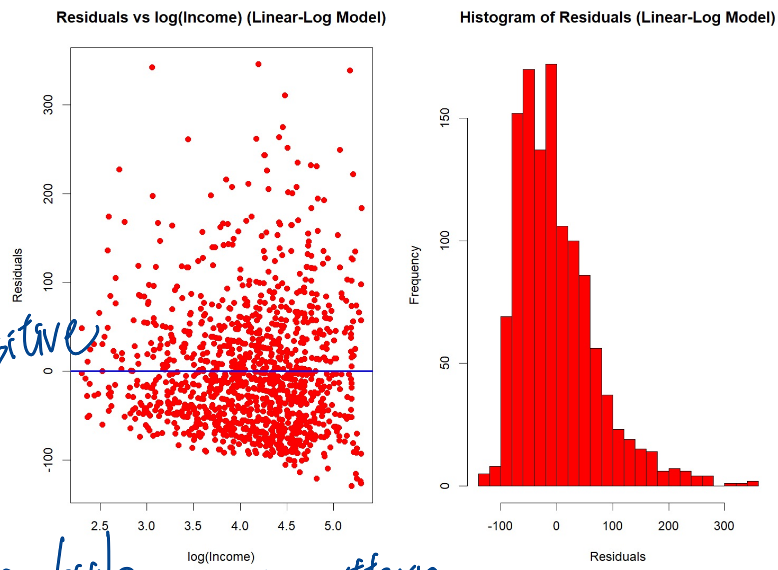
	income	food_predicted	elasticity	lower_95_ci	upper_95_ci
1	19	88.90	0.2496	0.1784	0.3208
2	65	116.19	0.1910	0.1365	0.2454
3	160	136.17	0.1629	0.1165	0.2094

linear-log : elasticity ↓ when income ↑

linear : elasticity ↑, income ↑

log-log : fixed

(j)



positive

a little spray partern

X-squard : 628.07, df=2

p-value < 2.2e-16

greater than 5.99 ⇒ reject H_0.

The linear-log regression errors are normal

(k) The log-log model assumes constant income elasticity across all income levels, with the most random residuals and the least skewness. Based on this, it's a good choice.