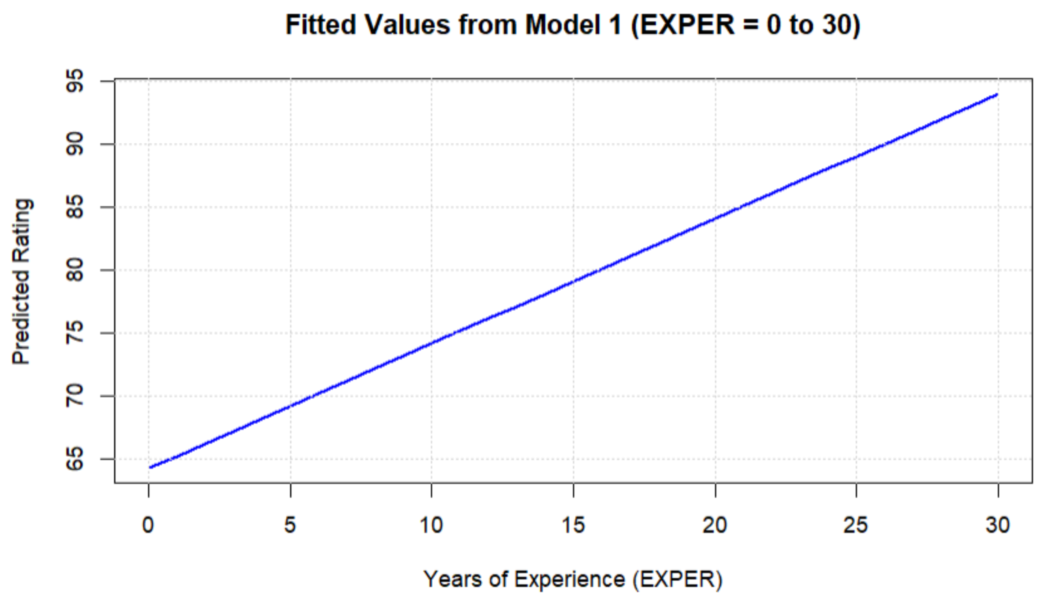
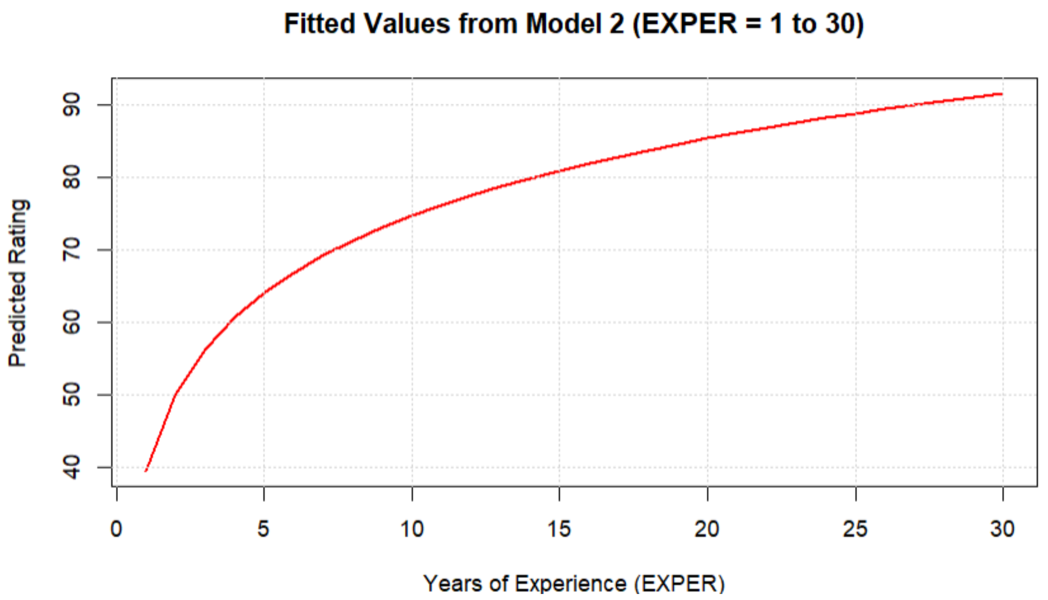
313707047 林均桐

CH04.Q4

(a)



(b)



Model 2 使用 ln(EXPER)，而 ln(0)沒有定義，所以沒有經驗的 4 名藝術家不包含在這個模型中。

(c) 邊際效應就是斜率，對任意一點都是固定的 **0.990**，因為是線性模型所以(i)

10 年經驗、(ii) 20 年經驗下的邊際效應都是 **0.990**。

(d) 當 EXPER=10，邊際效應= $\frac{15.312}{10}$ =1.5312

當 EXPER=20，邊際效應= $\frac{15.312}{20}$ =0.7656

(e) Model 1 R^2=0.3793

Model 1 (有經驗的) R^2=0.4858

Model 2 R^2=0.6414

所以 Model 2 對資料的擬合最好

(f) Model 2 較為合理，因為它使用對數形式，代表隨著工作經驗增加，評分上升的速度會遞減。

這反映「邊際報酬遞減」的經濟直觀。

相反地，Model 1 是線性模型，表示無論經驗多寡，每增加一年經驗對評分的提升都是固定的，這通常不符合實際情況。

CH04.Q28

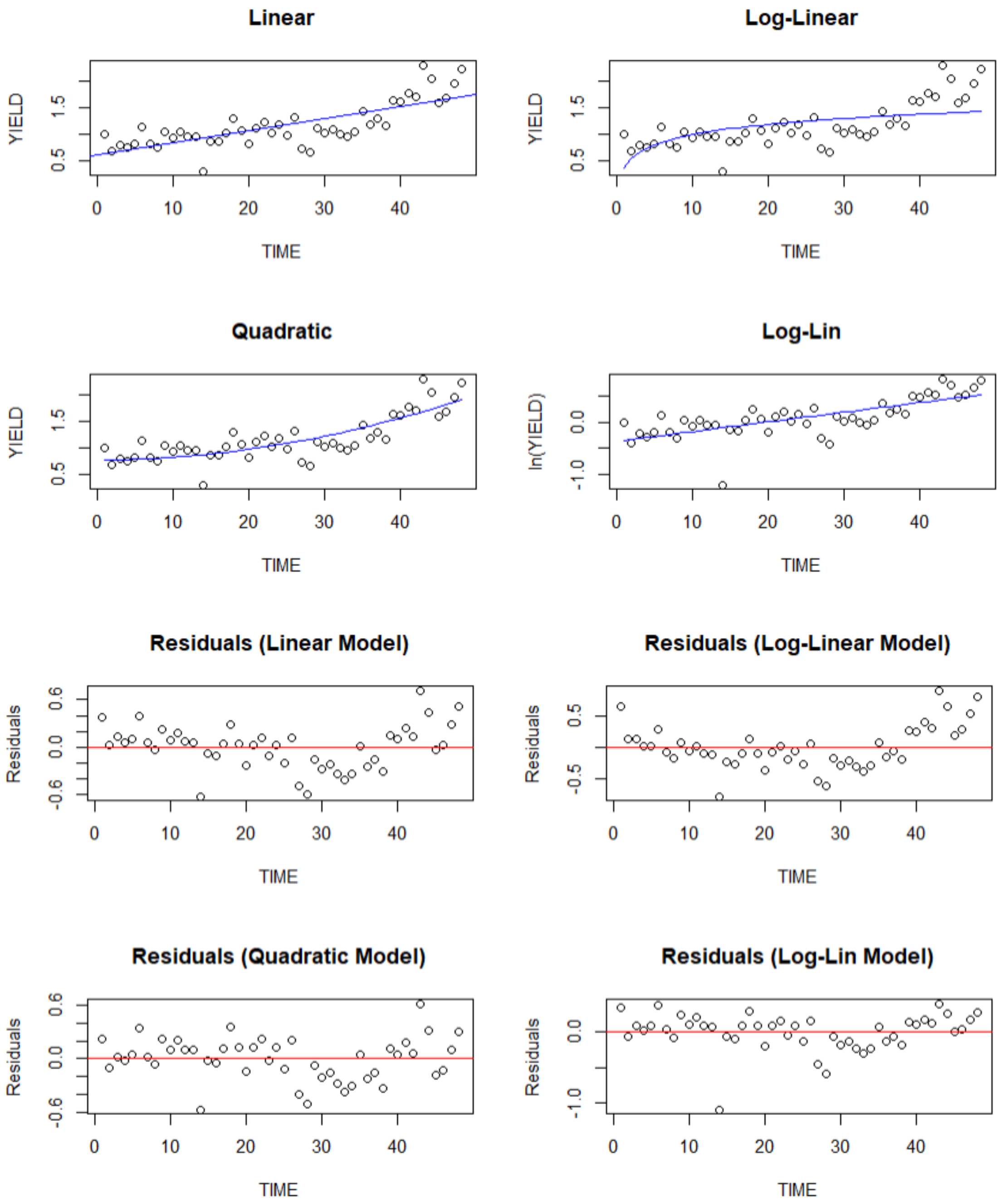
(a)

線性模型: $YIELD=0.6032+0.0231\times TIME$

log-linear 模型: $YIELD=0.351+0.279\times \ln(TIME)$

二次函數模型: $YIELD=0.7737+0.000499\times TIME^2$

log-lin 模型: $\ln(YIELD)=-0.3639+0.0186\times TIME$



```
        Jarque Bera Test

data:  resid(model1)
X-squared = 0.13257, df = 2, p-value = 0.9359


data:  resid(model2)
X-squared = 2.7629, df = 2, p-value = 0.2512


data:  resid(model3)
X-squared = 0.32406, df = 2, p-value = 0.8504


data:  resid(model4)
X-squared = 83.874, df = 2, p-value < 2.2e-16
```

線性模型 Multiple R-squared: 0.5778

log-linear 模型 Multiple R-squared: 0.3386

二次函數模型 Multiple R-squared: 0.689

log-lin 模型 Multiple R-squared: 0.5074

最適合模型：

二次函數模型的R^2最高，且殘差分佈較為隨機，沒有明顯趨勢或漏斗形狀，但若是選擇根據 Jarque Bera Test 則會因為 P 值是最高的代表可能較符合常態分布

(b)

在我選擇的二次函數模型中，時間相關變數的係數代表 $TIME^2$的係數 0.000499 為正，隨著時間的推移，產量的成長有加速的趨勢，且隨著時間平方增加，YIELD 會以遞增速率上升，係數顯示時間越久，產量成長速度也越快。

(c)

```
>     # (1) 計算學生化殘差
>     stud_res <- rstudent(model3)
>     which(abs(stud_res) > 2)   # 找出大於 |2| 的異常觀測值
14 28 43
14 28 43
```

```
>     # (2) 槓桿值 (leverage)
>     lev <- hatvalues(model3)
>     # 通常判斷標準：2p/n，其中 p = 參數數量，n = 樣本數
>     p <- length(coef(model3))
>     n <- nrow(wa_wheat)
>     threshold_lev <- (2 * p) / n
>     which(lev > threshold_lev)
45 46 47 48
45 46 47 48
```

```
>     # (3) DFBETAS：判斷哪些觀測值對係數影響大
>     dfb <- dfbetas(model3)
>     # 判斷標準：絕對值 > 1
>     which(abs(dfb[,2]) > 1)   # 看 TIME² 的影響
named integer(0)
```

```
>     # (4) DFFITS：判斷對整體擬合影響大的觀測值
>     dff <- dffits(model3)
>     threshold_dffits <- 2 * sqrt(p / n)
>     which(abs(dff) > threshold_dffits)
14 43 48
14 43 48
```

(d)

```
    print(pred_1997)
      fit      lwr      upr
 1.875113 1.367898 2.382328
```

```
>     print(actual_1997)
[1] 2.2318
>     # 看 actual_1997 是否落在 pred_1997 的 Lwr 與 Upr 區間內
```

預測區間為 (1.3679 ~ 2.3823) ，2.2318 有落在這個區間內

CH04.Q29

(a)

Food 樣本平均數 114.4431

Food 樣本中位數 99.8

Food 最小值 9.63

Food 最大值 476.67

Food 標準差 72.6575

income 樣本平均數 72.14264

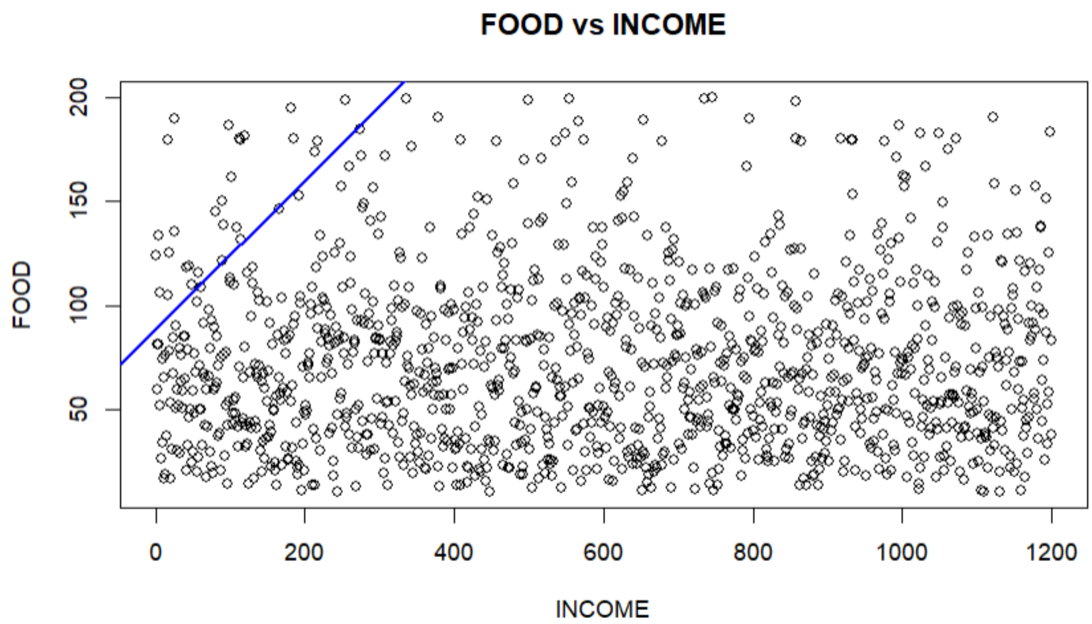
income 樣本中位數 65.29

income 最小值 10

income 最大值 200

income 標準差 41.65228

(b)

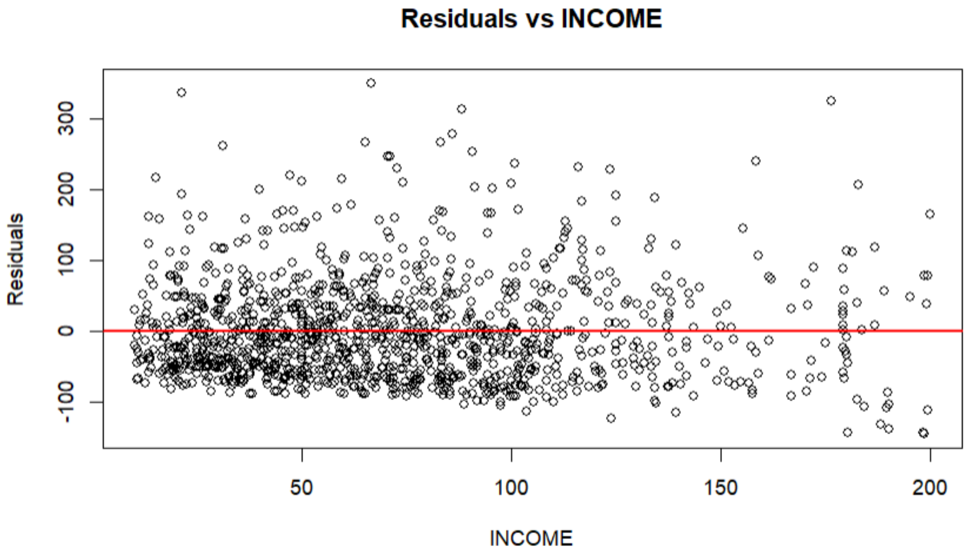


FOOD =88.566+0.359×INCOME

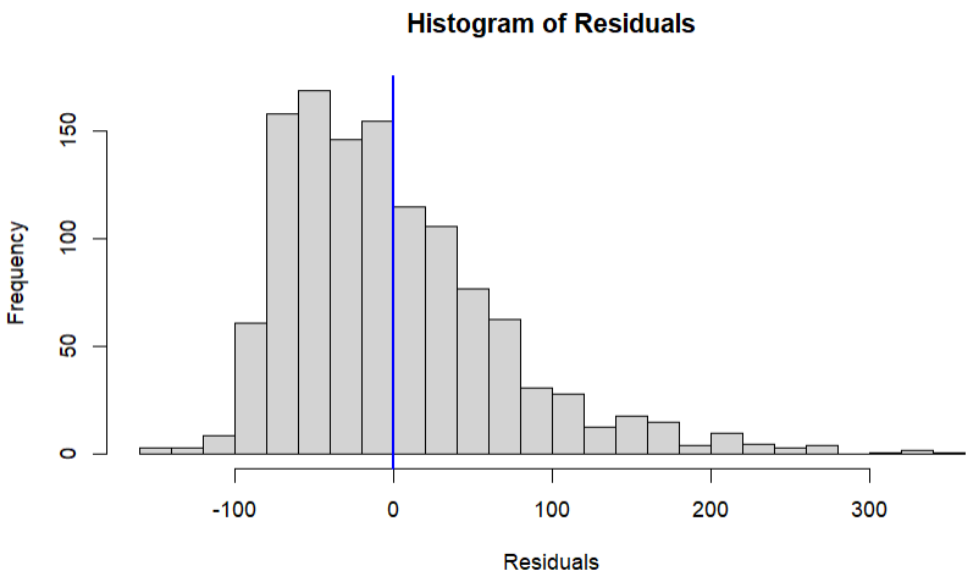
95%信賴區間 [0.2619215 , 0.455452]

整個區間都在 0 以上，表示收入對 food 支出有顯著正向影響，應該相對精確

(c)



當 INCOME 很低時，殘差分布比較集中，而當 INCOME 增加時，殘差的變異性變大，有較多大殘差在高收入區域出現，表示模型在高收入區域預測不穩定。



左邊殘差分佈較集中，右邊有較長的尾巴，不是對稱的鐘型分佈，因此不符合常態性假設。

誤差項呈常態分佈比較重要。誤差項要有平均數為 0、常態分佈、且同質性這樣使用 t 檢定和 F 檢定、信賴區間才有效，自變數和應變數本身不一定需要常態分佈，因為最小平方法估計沒有這個限制。

```
        Jarque Bera Test

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

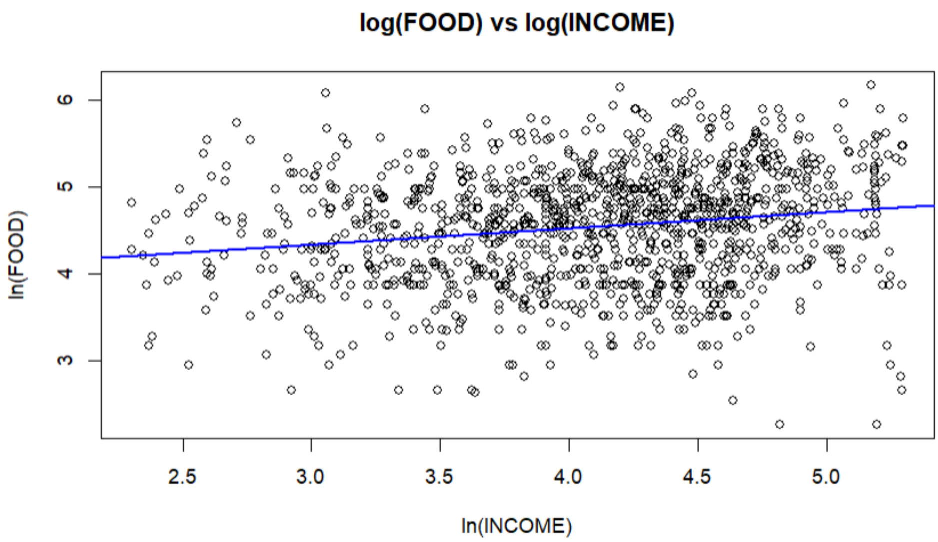
因 p 值（遠小於 0.05），非常態分佈

(d)

INCOME	Point_Estimate	Lower_95_CI	Upper_95_CI
19	0.07145038	0.05217475	0.09072601
65	0.20838756	0.15216951	0.26460562
160	0.39319883	0.28712305	0.49927462

(e)

ln(FOOD)=3.779+0.186ln(INCOME)



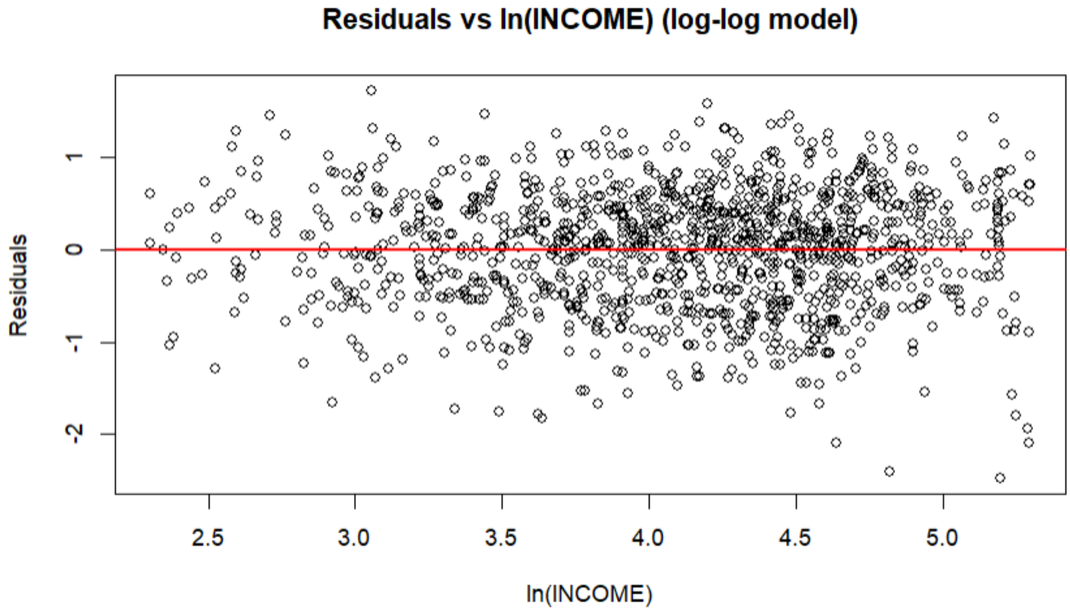
有更明顯，雖然分散度還是存在但整體趨勢比較穩定、有正向關係，因為 log-log 模型較符合邊際遞減和收入彈性的概念。

(f)

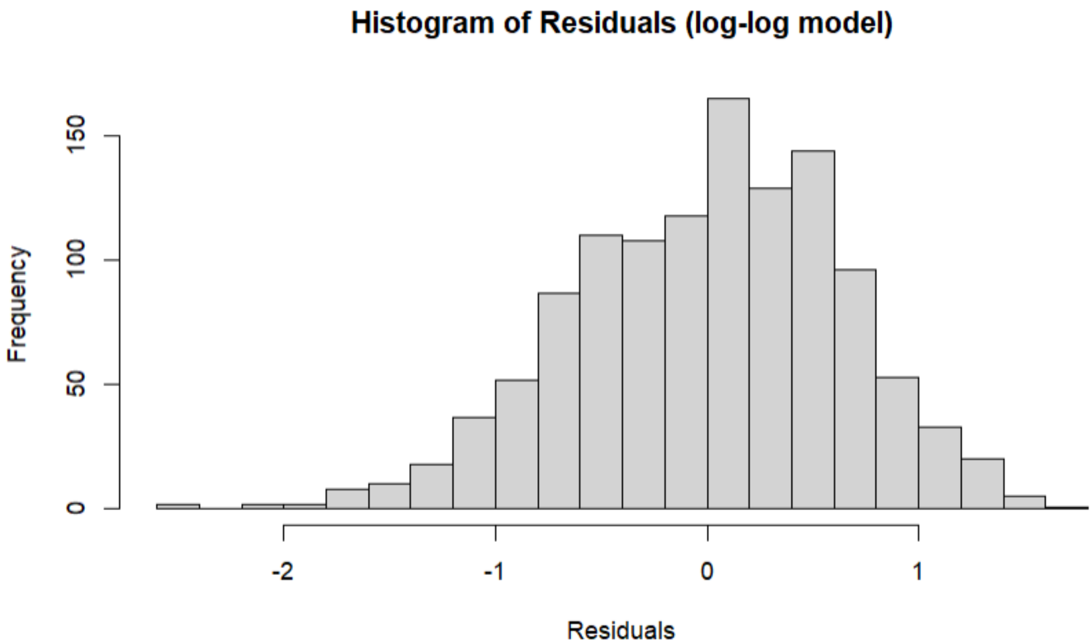
95%信賴區間 [0.129 , 0.243]

和 d 不相似，線性模型彈性隨 INCOME 改變，而 log-log 模型彈性是固定常數

(g)



殘差分佈在 ln(INCOME)範圍內大部分隨機分佈



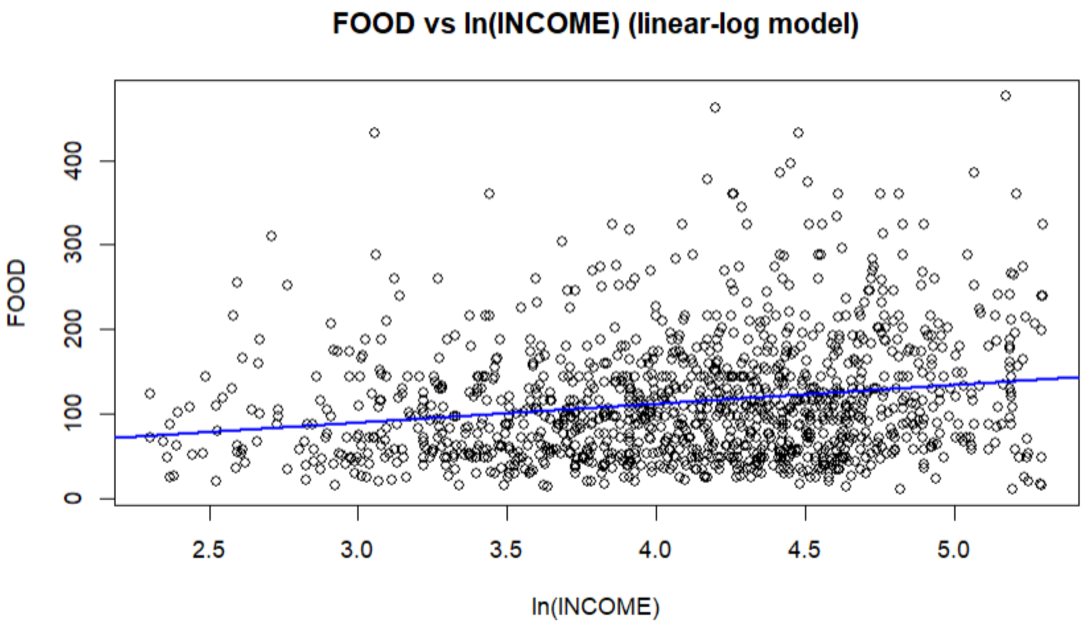

```
        Jarque Bera Test

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

p-value = 2.436e-06（遠小於 0.05），非常態分佈

(h)

FOOD=23.6+22.2·ln(INCOME)+e



將 b、e、h 做比較後發現 e 的模型關係比較清晰

```
# 比較 R²
  cat("R² for linear model:", summary(model_b)$r.squared, "\n")
for linear model: 0.0422812
  cat("R² for log-log model:", summary(model_e)$r.squared, "\n")
for log-log model: 0.03322915
  cat("R² for linear-log model:", summary(model_h)$r.squared, "\n")
for linear-log model: 0.03799984
```

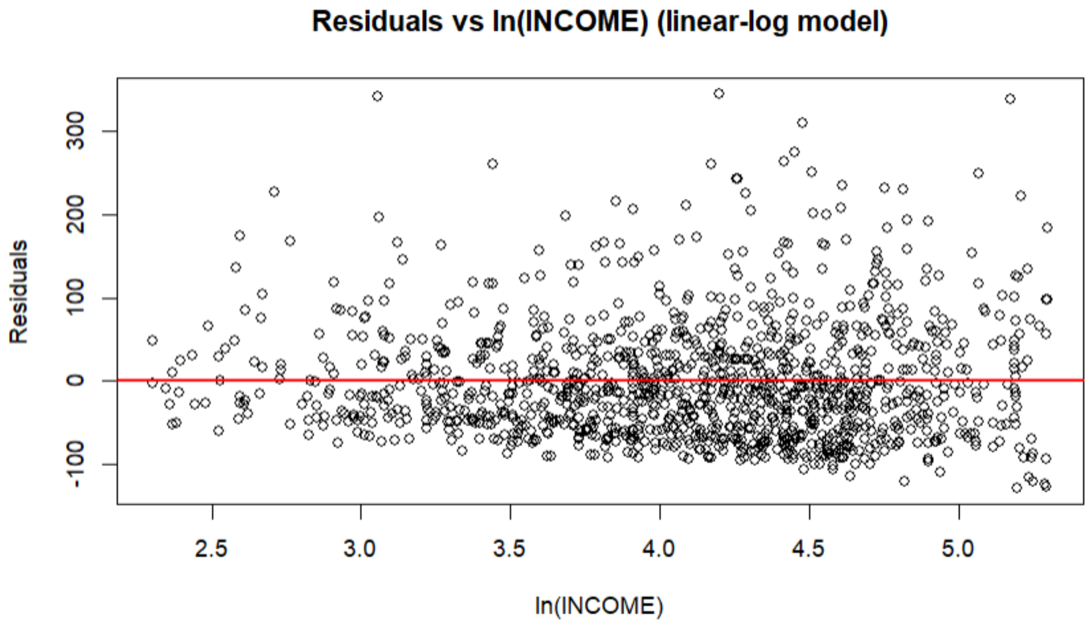
擬合度最好的是 (b) 線性模型

(i)

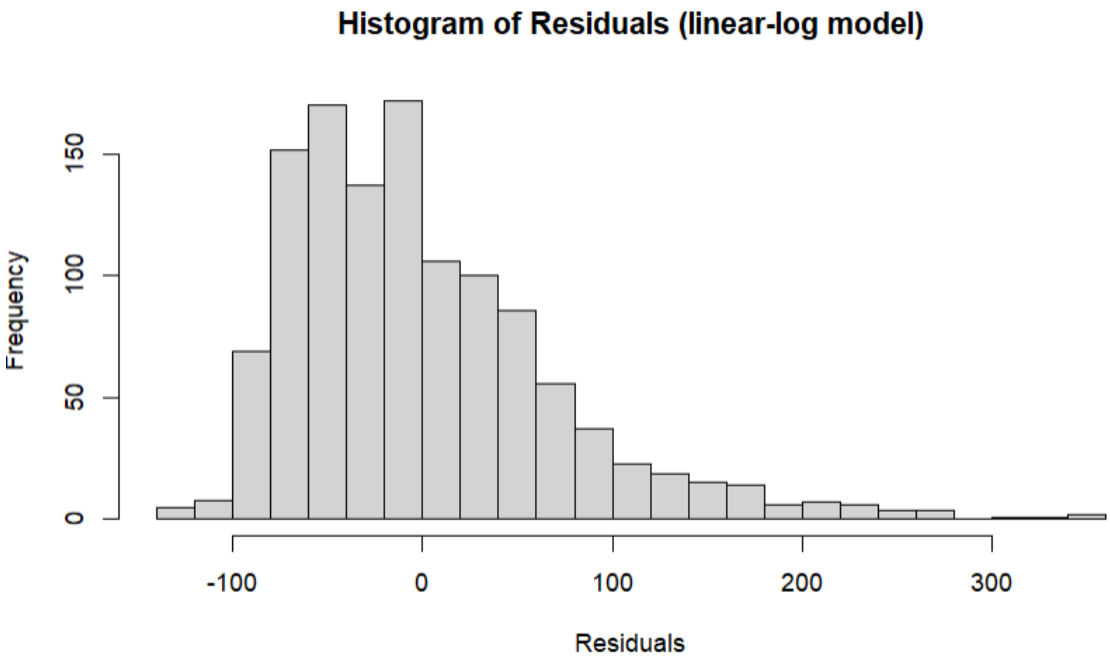
INCOME	Elasticity	Lower_95_CI	Upper_95_CI
19	0.2495828	0.1784009	0.3207648
65	0.1909624	0.1364992	0.2454256
160	0.1629349	0.1164652	0.2094046

linear-log 模型與線性模型的結果相似，都是可變的彈性，與 log-log 模型結果在中等收入時相近，但 log-log 模型是一個固定值而 linear-log 模型是可變的。

(j)



沒有明顯系統性趨勢，殘差在 ln(INCOME)不同區間的分佈相對較隨機，隨著 ln(INCOME)增加，殘差似乎略微變大，而在高 ln(INCOME)區域殘差值偏大。




```
        Jarque Bera Test

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

p-value < 2.2e-16，所以非常態分佈。

(k)

我偏好使用 log-log 模型，因 log-log 中的係數本身就是收入彈性，直接代表收入增加 1%時，食物支出增加多少百分比，雖然 log-log 模型的 R^2 並非最高，但殘差分佈較穩定且較不受異質性影響，且擁有固定的彈性係數，而當資料偏態或變異性大時，取 log 能減低極端值影響。