Name: Nguyen Quoc Nhan

ID:413707009

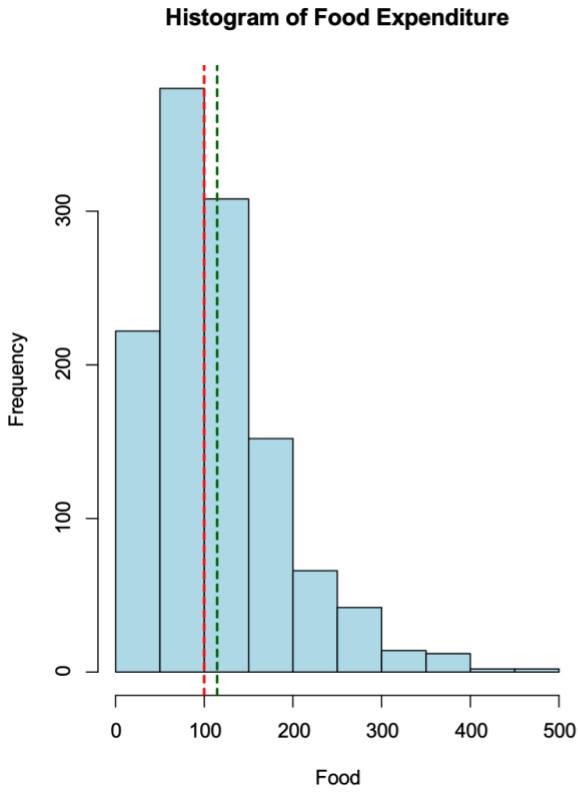
HW0317

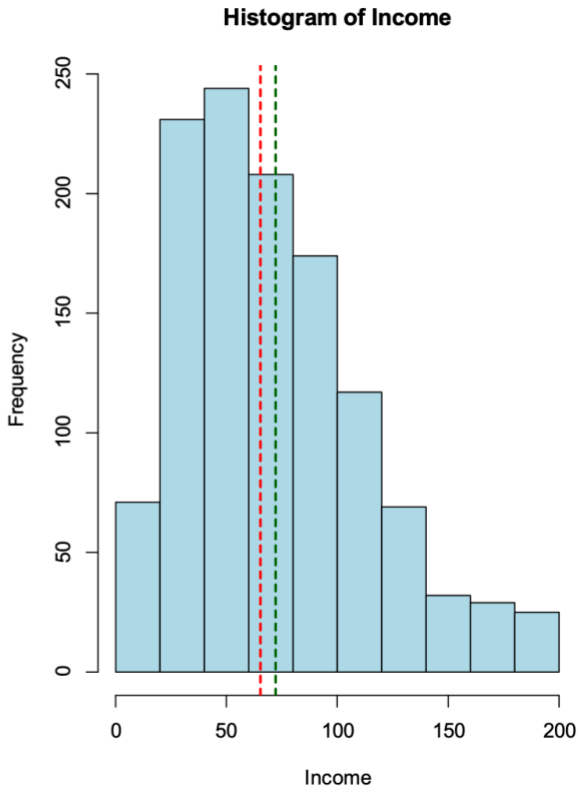
Question 29

a.

```
> print(food_summary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   9.63   57.78   99.80  114.44  145.00  476.67
> cat("Standard Deviation for food:", food_sd, "\n\n")
Standard Deviation for food: 72.6575
```

```
> print(income_summary)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  10.00   40.00   65.29   72.14   96.79  200.00
> cat("Standard Deviation for income:", income_sd, "\n\n")
Standard Deviation for income: 41.65228
```





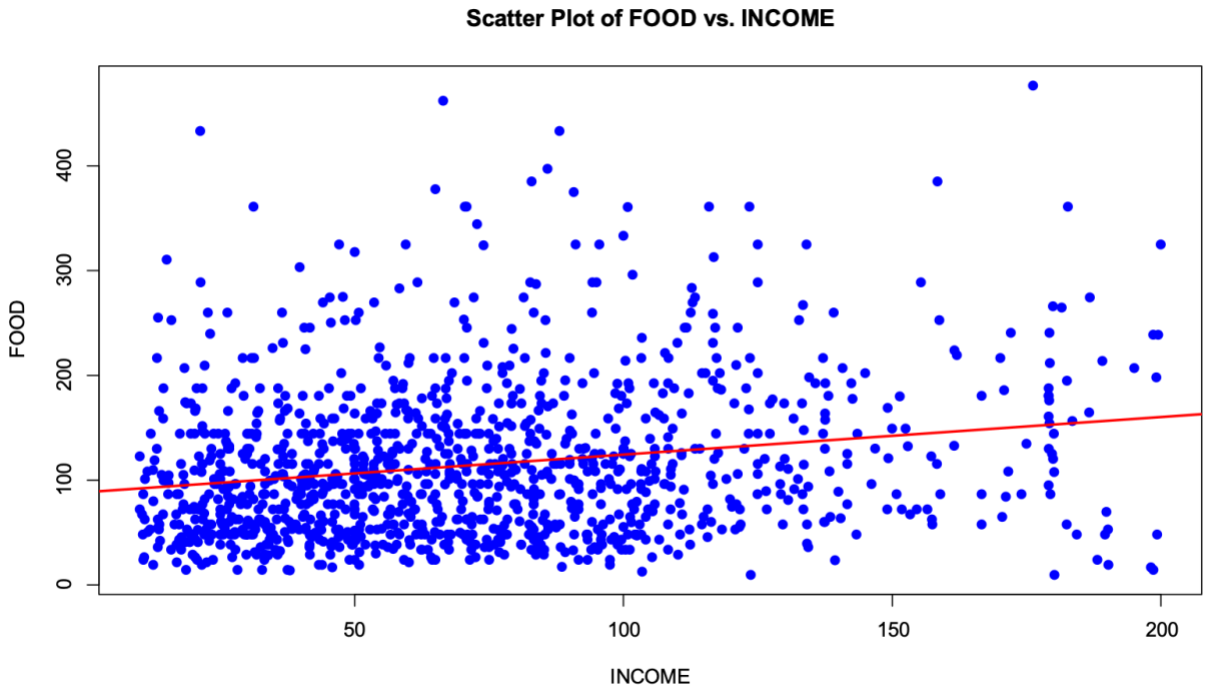

```
    Jarque Bera Test

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

```
    Jarque Bera Test

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

The histograms are not asymmetrical and bell shape. When we check the Jarque Bera Test, the p-value of food and income is smaller than 5%. It rejects the null hypothesis of normality.

b.

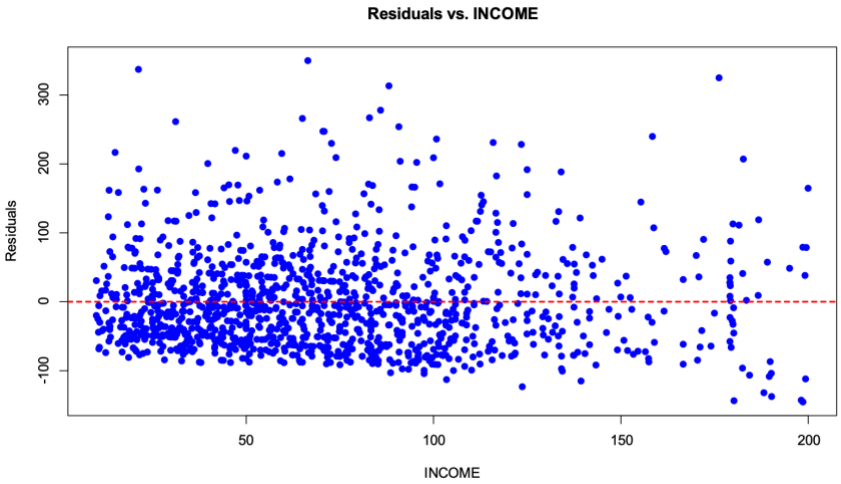
```
lm(formula = food ~ income, data = cex5_small)

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
```

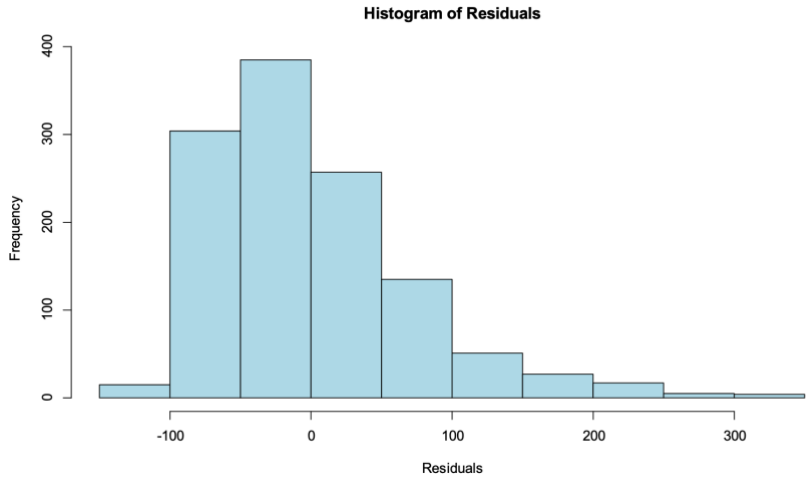



```
> print(confint_beta)
              2.5 %    97.5 %
income 0.2619215 0.455452
```

Beta2 at 95% confidence intervals ranging from 0.2619215 to 0.455452. We cannot estimate the effect of changing income on average FOOD precisely.

c.





The pattern between residuals and INCOME is skewness.

The Jarque Bera Test statistics are 642.19 with p-value < 0.01. It reject null hypothesis of normality.

It is more important that the error term e (or residuals) be normally distributed because many of the inferential procedures in regression (like t-tests for coefficients) rely on the normality assumption of the error term. The normality of FOOD or INCOME is not required for valid inference; what matters is that the residuals capture the random error properly and follow a normal distribution, which ensures that the statistical tests are reliable.