NATIONAL ECONOMICS UNIVERSITY

FACULTY OF MATHEMATICAL ECONOMICS

GROUP PROJECT

SUBJECT: ECONOMETRICS

Topic: Identify factors that affect student expenses

Class:	DSEB K61
Group:	4
Teacher:	Assoc. Prof.Nguyen Thi Minh
Member:	Nguyen Thuy Linh Dao Thi Hong Nhung Bui Thi Mai Luong Nguyen Thi Hoai Linh

Ha Noi, 2021

Contents

A. Introduction

The monthly expenses are issues that everyone is concerned, especially students, because most of them have a new life away from home. They have to live independently and take care of their monthly expenses. Because they of inexperience, the spending is not reasonable. To help them find a solution to balance their monthly expenses, our group has chosen the research topic: **" Identify factors that affect student expenses"**

First, our group collected data and analyzed the basic values such as: Min, Max, Mean, Kurtosis, etc.

Second, we ran the data to select variables through the following methods: checking multicollinearity, writing linear regression equations for all variables, testing the selection of variables through the "ols_step_best_subset" function.

Then, because the regression equation after selecting the variable did not satisfy the tests for normal distribution and constant variance, we continued to transform the model.

After selecting the appropriate transform model, we checked the influence point and outliers. Next, we split data into two parts (70 train – 30 test) and test the transform model in 70 % of the data to check if the model had a big difference. Finally, we checked the PRESS coefficient, compared MSE and Residual mean square.

B. Theoretical basis

I. Selection of variables for the model

There are many factors that affect a student's monthly expenses that we can see such as entertainment, meals, travel expenses, etc. However, our group has selected 6 main factors that greatly influence student spending.

Include:

• Family allowance for students (Sup)

• Income from part-time work (Income)

• Their accommodation is represented by two numbers 0 and 1 (0 means no house in Hanoi, have to rent a house and 1 means have a house in Hanoi) (Home)

• Gender with male is 1 and female is 0 (GEN)

• Monthly food/drink expenses (Eating)

• Scholarship where 0 is no and 1 is yes (Scholarship)

II. Description of variables

Variable type	Sign	Meaning	Sign expectation
Dependent variable	**Expense**	Average total monthly spending of students measured in Vietnamese dong (unit: thousand VND).	
Independent variable	**Sup**	Support: Monthly family support measured in Vietnamese dong (unit: thousand VND).	+
Independent variable	**Income**	Income: Income of students with part-time jobs measured in Vietnam dong (unit: thousand VND).	+
Independent variable	**Home**	Home: is represented by two numbers 0 and 1 (0 means no house in Hanoi, have to rent a house and 1 means have a house in Hanoi)	-
Independent variable	**Gen**	Gender: Male – 1, Female: 0	+

Independent variable	**Eating**	Eating: Monthly food/drink expenses measured in Vietnam dong (unit: thousand VND).	+
Independent variable	**Scholarship**	Scholarship: student who has scholarship is 1 and 0 is no	-

C. Data analysis

I. Descriptive statistics

	Expense	sup	Income	Home	gen	Eating	Scholarchip
Mean	2907	2761	3526	0.6591	0.4924	1424	0.4848
Median	2850	2500	3500	1	0	1430	0
Mode	2500	2500	2500	1	0	1499	0
1st Qu.	2350	1500	2500	0	0	1117	0
3rd Qu.	3334	3500	4500	1	1	1739	1
Min	1120	1000	800	0	0	838	0
Max	7850	6000	8000	1	1	1993	1
Standard Error	76.668	109.209	122.91	0.0414	0.044	30.5287	0.0437

S.Dev	834.893	1254.71	1412.11	0.4758	0.502	350.75	0.5017
Kurtosis	8.6117	-0.2753	0.4925	-1.563	-2.03	-1.255	-2.0272
Skew	1.8017	0.6379	0.7235	-0.679	0.031	0.0138	0.0613

II. Regression

1. Checking for multicollinearity

- Excel

	expense	sup	income	home	gen	eating	scholarship
expense	1						
sup	0.1206701	1					
income	0.5391629	-0.009846	1				
home	0.0397575	0.0963128	0.2206703	1			
gen	0.0978639	0.1569103	-0.009859	-0.021597	1		
eating	0.2745019	0.0776652	0.0347148	-0.094257	0.1466784	1	
scholarship	0.0533155	0.053519	0.0469656	0.1493219	-0.007931	0.0730053	1

Correlation matrix of model

- R_code

```
model<-lm(expense~.,data)
library(faraway)
vif(model)
      sup    income      home       gen    eating scholarship
 1.042243  1.056103  1.102548  1.045955  1.047629    1.034170
```

Conclusion: The correlation coefficients between the variables in the excel have absolute values less than 0.8 and retested by R has values less than 3, so there is no correl ation between the variables.

- The model does not suffer from multicollinearity.

2. Write the linear regression equation

```
> model <-lm(expense~.,data=data)
> summary(model)

Call:
lm(formula = expense ~ ., data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1432.1  -407.2   -51.0   288.6  3324.5 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  816.90861  312.83512   2.611  0.01012 *  
sup            0.06779    0.04806   1.410  0.16089    
income         0.32322    0.04299   7.519 9.34e-12 ***
home        -128.72526  130.34883  -0.988  0.32528    
gen           94.91057  120.37514   0.788  0.43192    
eating         0.55656    0.17237   3.229  0.00159 ** 
scholarship   17.50146  119.73630   0.146  0.88403    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 676.1 on 125 degrees of freedom
Multiple R-squared:  0.3743,	Adjusted R-squared:  0.3443 
F-statistic: 12.46 on 6 and 125 DF,  p-value: 5.658e-11
```

- Expense = 816.91 + 0.07*Sup + 0.32*Income - 128.72*Home + 94.81*Gen + 0.56*Eating +17.5*Scholarship
- The model has some variable such as: Home, Gen, and Scholarship is insignificant because the P_value is greater than 0.17.

  Prediction: We will choose 3 variables: Sup, Income and Eating to rebuild the regression model.

  2.3. Verify the selection of variables

```
> library(olsrr)
> ols_step_best_subset(model_0)
               Best Subsets Regression
-----------------------------------------------------
Model Index    Predictors
-----------------------------------------------------
     1         income
     2         income eating
     3         sup income eating
     4         sup income home eating
     5         sup income home gen eating
     6         sup income home gen eating scholarship
-----------------------------------------------------

                                                      Subsets Regression Summary
-----------------------------------------------------------------------------------------------------------------------------------------
                       Adj.        Pred
Model    R-Square    R-Square    R-Square     C(p)        AIC         SBIC         SBC           MSEP            FPE           HSP       APC
-----------------------------------------------------------------------------------------------------------------------------------------
  1        0.2890      0.2835      0.2419    14.0458    2110.5813    1735.6886    2119.2297    65922602.1381    506979.6239    3871.4168    0.7329
  2        0.3557      0.3457      0.3012     2.7148    2099.5727    1725.1247    2111.1039    60202463.6911    466415.8485    3562.8988    0.6742
  3        0.3662      0.3513      0.2996     2.6314    2099.4186    1725.1544    2113.8327    59694388.7783    465876.8991    3560.4302    0.6735
  4        0.3711      0.3513      0.2947     3.6367    2100.3776    1726.2792    2117.6744    59695501.5778    469282.1158    3588.5320    0.6784
  5        0.3742      0.3494      0.2878     5.0214    2101.7296    1727.7934    2121.9092    59878366.7418    474125.7109    3628.0924    0.6854
  6        0.3743      0.3443      0.2762     7.0000    2103.7070    1729.8850    2126.7694    60350941.7592    481299.6332    3685.9800    0.6958
-----------------------------------------------------------------------------------------------------------------------------------------
AIC: Akaike Information Criteria
 SBIC: Sawa's Bayesian Information Criteria
 SBC: Schwarz Bayesian Criteria
 MSEP: Estimated error of prediction, assuming multivariate normality
 FPE: Final Prediction Error
 HSP: Hocking's Sp
 APC: Amemiya Prediction Criteria
```

Table 1

Stepwise presents the results of 6 models evaluated as the most optimal for the dependent variable (expense). Observe that model 6 of this method is the original model.

- The first column is the STT of the best-fit models.
- The third column is adjusted R-Square. Looking at Table 1, model 2,3,4,5 has adjusted R-Square larger than model 6 (original model).
- The 5th column is the value C(p). A simple and complete model should be one with as low a C(p) value as possible. In Table 1, we see that there are 2 models with the smallest C(p) values, model 2 (2.7148) and model 3 (2.6314).

⇨ Through the stepwise method, there are two optimal models:

Model 2: Expense ~ $\beta_0 + \beta_1$*income + β_2*eating + ε

Model 3: Expense ~ $\beta_0 + \beta_1$*income + β_2*eating + β_3*sup + ε

- Compare model 2 with model 3

Model: expense ~	~ $\beta_0 + \beta_1$*income + β_2*eating + ε	~ $\beta_0 + \beta_1$*income + β_2*eating + β_3*sup + ε	Note
Vif	Income 1.001209, Eating 1.001209	Sup 1.0058, Income 1.0013, Eating 1.007	Two model not multicollinearity
Press	63809753	63959698	The model with the lowest PRESS index is better.

⇨ Choose the model 2.

3. Checking adequacy model 2

Expense ~ $\beta_0 + \beta_1$*income + β_2*eating + ε

3.1. Eliminate insignificant variables and linear regression

```
> model = lm(expense ~ income + eating, data )
> summary(model)
```

```
Call:
lm(formula = expense ~ income + eating, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1377.7  -378.1   -82.4   327.9  3274.3 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 928.61833  283.11422   3.280 0.001335 ** 
income        0.31253    0.04181   7.475 1.04e-11 ***
eating        0.61525    0.16832   3.655 0.000373 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 675.3 on 129 degrees of freedom
Multiple R-squared:  0.3557,	Adjusted R-squared:  0.3457 
F-statistic: 35.61 on 2 and 129 DF,  p-value: 4.842e-13
```

- Conclusion: After removing insignificant variables, the output

  P_value is satisfied and the linear equation has the form:

  Expense = 928.62 + 0.31253 * Income + 0.61525 * Eating

3.2. Checking for multicollinearity

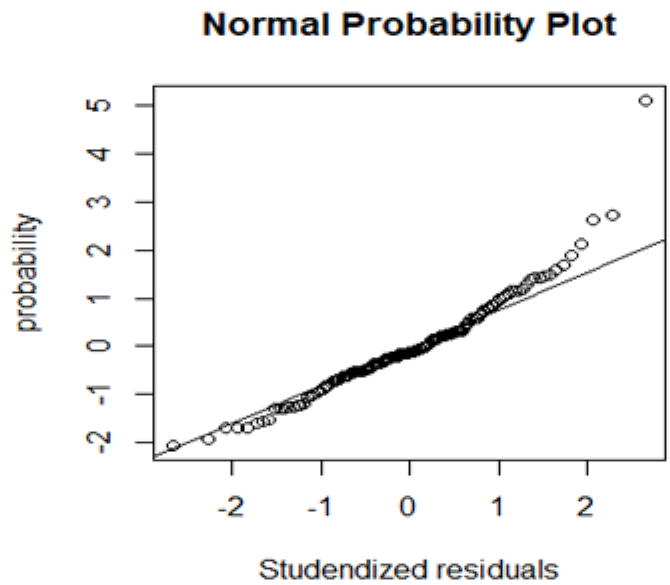
```
> vif(model)

      income   eating 
      1.001209 1.001209 
```

Conclusion: 2 independent variables are not in the case of multicollinearity.

3.3. Plot the Normal Probability Plot and check with Jarquebera test.

a) Plot the Normal Probability Plot

```
> a = rstandard(model)
> qqnorm(a, ylab = 'probability', xlab = 'Studendized residuals', main = 
"Normal Probability Plot" )
> qqline(a)
```



Conclusion: The graph is in the form of a light-tailed distribution. This suggests that may be the dependent variable "Expense" is not normal distribution.

b) Check with Jarquebera test.

Hypothesis H_0: Expenses are normally distributed

H_1: Expenses are not normally distributed

```
> library(tseries)
> b = model$residuals
> jarque.bera.test(b)

Jarque Bera Test

data:  b
X-squared = 95.102, df = 2, p-value < 2.2e-16
```

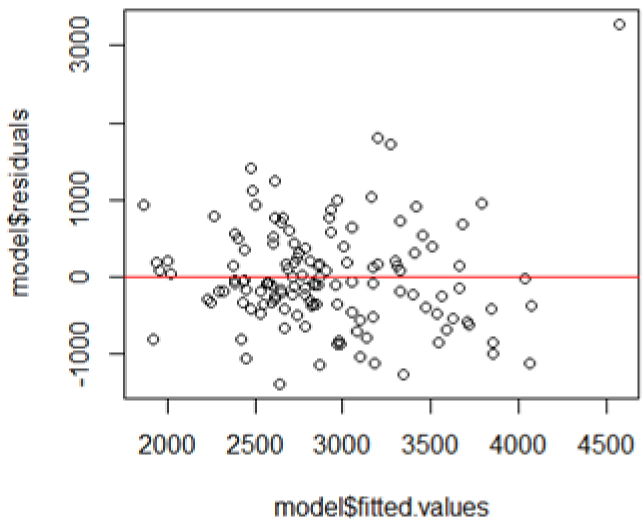
Conclusion: P_value < 0.05 => Reject H_0 => 'Expense' is not normally distributed.

3.4. Checking if independent variables have var constant or not?

- Checking by graphing

```
> plot(model$fitted.values, model$residuals)
```

```
> abline(h = 0, col = 'red')
```



Conclusion: Maybe independent variables have var non-constant => Transform linear equation

- Checking by bptest

```
> library(lmtest)
> bptest(model)
Studentized Breusch-Pagan test

data:  model
BP = 16.534, df = 2, p-value = 0.0002569
```

Conclusion: P_value < 0.05, so we reject the hypothesis that the H_0 error term definitely has a var non-constant.

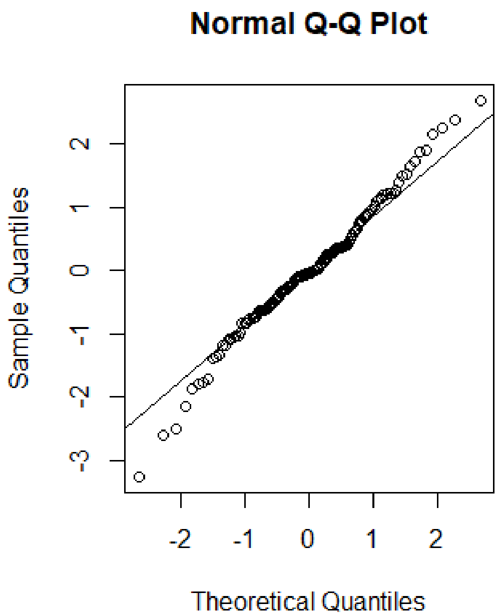
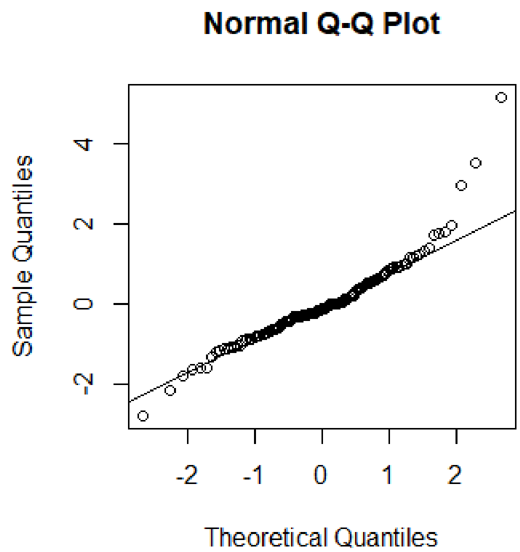
4. Transform function and checking model adequacy

4.1 Checking model

Test function: model_1 = lm(log(expense) ~ sqrt(income) + eating, data)

Test function: model_2 = lm(log(expense) ~ income + sqrt(eating), data)

	Model_1	Model_2

Regression	Log(expense) =6.962 + 0.01119*sqrt(income) + 2.283 eating	Log(expense) = 6.984+ $9.653*10^{-5}$ income+0.016*sqrt(eating)
Normal Plot		
Jarque Bera Test	P_value = 0.1863 > 0.05 Satisfy	P_value = 0.1919 > 0.05 Satisfy
R-square	33.25%	34.21%
Adjust R square	32.21%	33.19%
PRESS	6.826413	4.249736
Vif	Sqrt(income) Eating 1.001059 1.001059 No multicollinearity	Income Sqrt(eating) 1.00115 1.00115 No multicollinearity

Test var con stant – bptes t	P_values = 0.2119 > 0.05 Model has no variable variance	P_values = 0.1074 > 0.05 Model has no variable variance

Table 2

Conclusion: Both models are satisfactory but model 2 has a smaller Press and has a highe r adjust R square -> So we will choose model 2

Log(expense) = 6.984 + 9.653*10^{-5} income + 0.016*sqrt(eating)

4.2 Regression coefficient test

We conduct the regression coefficient β_i; i ∈[2,3] with significance level = 5% by

P-value method.

- Hypothesis: H_0: $\beta_2 = 0$

  H_1: $\beta_2 \neq 0$

According to the estimated results of model 2, we have: P_value = 9.653*10^{-5}< 0.05.

Therefore, we reject the hypothesis H0 that the extra monthly income actually affects the average monthly expenditure of students with 95% confidence.

- Hypothesis: H_0: $\beta_3 = 0$

  H_1: $\beta_3 \neq 0$

According to the estimated results of model 2, we have: P_value = 1.638*10^{-5}< 0.05.

Therefore, we reject the hypothesis H_0 that the extra monthly income actually affects the average monthly expenditure of students with 95% confidence.

4.3 Checking the fit of the model

- Hypothesis: H_0: $R^2 = 0$

H_1: $R^2 \neq 0$

According to the estimated results of model 3, we have: P_value = $1.877*10^{-12} < 0.05$

⇨ We reject the null hypothesis H_0, the found model fits with 95% confidence.

On the other hand: R^2 = 34.21% explains that the independent variables explain 34.21% of the change of the dependent variable, the remaining 65.79% because other variables have not been included in the model and have not been used

4.4 Meaning of regression coefficient

- B_2 hat: under the condition that other factors remain constant, if a student's monthly extra income increases by 1,000 VND, the average student spending in that month will increase by $9{,}653*10^{-5}$ thousand VND.

- B_3 hat: under the condition that other factors remain constant, if the student's food expenditure increases by 1,000 dong, the average student's spending in that month will increase by $1{,}638*10^{-2}$ thousand dong.

4.5 Data splitting

a) Model testing in 70% of data

- Splitting data: 30 – 70

```
library(caTools)
sample=sample.split(data$expense,SplitRatio = 0.7)
train = subset(data,sample==TRUE)
test1=subset(data,sample==FALSE)
```

- Validation

- PRESS

```
> library(MPV)
> ml3   = lm(log(expense) ~ income + sqrt(eating), train)
> op3 = summary(ml3)
> model = lm(log(expense) ~ income + sqrt(eating), data)
> PRESS3 = PRESS(ml3)
> PRESS4 = PRESS(model)
```

PRESS3 = 4.249736

PRESS4 = 4.249736

Conclusion: Press of the transform model has no difference.

- Compare average square prediction error(MSE) and residual mean square

```
> library(olsrr)
> predR2 = ols_pred_rsq(ml3)
> newy   = predict(ml3,test1)
> ntest1 = cbind(test1,newy)
> nres   = log(ntest1$expense) - ntest1$newy
> nMSE   = sum(nres^2)/nrow(test1)
> MSres3 = op3$sigma^2
> nMSE
> Msres3
```

Msres3 = 0.04296347 nMSE = 0.06180813

Conclusion: MSE and MSRres3 are not different. Therefore, the transform model is highly efficient.

b) Check influence point

```
2*sqrt(4/132) #compare abs DFFITS
[1] 0.3481553
2*sqrt(1/132) #compare abs DFBetas
[1] 0.1740777
#compare COV Ratio > 1+3p/n= (1+3*3/132) ở COV Ratio < 1-3p/n = 1-3*3/132
1+(3*3)/132
[1] 1.068182
1-(3*3)/132
[1] 0.9318182
```

```
> options(max.print=9999)
> influence.measures(model_2)
Influence measures of
	 lm(formula = log(expense) ~ income + sqrt(eating), data = data) :

      dfb.1_  dfb.incm  dfb.sq..     dffit cov.r   cook.d     hat inf
1   -4.93e-02  3.11e-04  4.01e-02 -9.86e-02 1.008 3.24e-03 0.00908
2    3.88e-02  6.48e-02 -6.64e-02 -1.01e-01 1.061 3.42e-03 0.04099
3   -4.20e-01  6.98e-01  2.48e-01  7.81e-01 0.986 1.96e-01 0.09360   *
4    1.78e-02  4.27e-02 -3.57e-02 -6.19e-02 1.058 1.29e-03 0.03486
5    4.32e-02  6.40e-04 -4.04e-02  5.64e-02 1.035 1.07e-03 0.01560
6    5.46e-02  2.16e-02 -7.03e-02 -8.91e-02 1.039 2.66e-03 0.02289
7   -7.14e-03  5.15e-03  4.93e-03 -1.03e-02 1.039 3.56e-05 0.01507
8    8.40e-02 -7.46e-02 -7.36e-02 -1.30e-01 1.031 5.61e-03 0.02346
9    1.53e-01 -1.27e-01 -1.32e-01 -2.09e-01 1.036 1.45e-02 0.03751
10  -1.61e-01  1.12e-01  1.13e-01 -2.27e-01 0.963 1.69e-02 0.01543
11  -7.96e-02  2.42e-02  6.67e-02 -1.01e-01 1.024 3.40e-03 0.01536
12  -2.01e-01  1.60e-01  1.32e-01 -3.08e-01 0.893 3.04e-02 0.01421   *
13  -1.64e-03  5.05e-03  3.87e-04  5.48e-03 1.082 1.01e-05 0.05404   *
14  -3.79e-03  7.54e-03  7.15e-04 -1.03e-02 1.041 3.57e-05 0.01657
15   1.12e-01 -1.10e-01 -6.28e-02  2.01e-01 0.965 1.33e-02 0.01286
16  -3.22e-02 -6.34e-04  3.10e-02 -3.83e-02 1.045 4.93e-04 0.02202
17   9.34e-02 -9.92e-02 -7.86e-02 -1.61e-01 1.016 8.66e-03 0.02084
```

➔ Conclusion: Influence Point

+ COV (point = values COV): 12 =0.893, 24 = 1.119 ,39 = 0.928, 51 = 0.901, 68 = 0.799 ,70 = 1.078

+ DF Fit: 3,21,40,51,64,68,99,101,103,119,122

+ DF Betas: 3,21,40,51,64,68,72,97,99,101,103,119,122

b.1) Check COV point

- All observations (12, 39, 51, 68) have the COVRATIO smaller than 1, so this observation degrades precision of estimation.
- All observations (24, 70) have COVRATIO greater than 1, so this observation tends to improve the precision.

b.2) DF fit and DF Betas

- Compare to model:

  + Model 1 is the transformed model in part 4.

  + Model 2 is the model which ran the regression with the data after deleting all DFFIT points and DF Betas points

- Code and result

```
> View(data_delete)
> model_delete = lm(log(expense) ~income +sqrt(eating),data_delete)
> summary(model_delete)

Call:
lm(formula = log(expense) ~ income + sqrt(eating), data = data_delete)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.55024 -0.10712 -0.00185  0.10555  0.38800 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.918e+00  1.410e-01  49.070  < 2e-16 ***
income       8.240e-05  1.313e-05   6.275 6.23e-09 ***
sqrt(eating) 1.917e-02  3.600e-03   5.326 4.98e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1773 on 116 degrees of freedom
Multiple R-squared:  0.3871,	Adjusted R-squared:  0.3765 
F-statistic: 36.63 on 2 and 116 DF,  p-value: 4.677e-13

> model=lm(log(expense) ~ income + sqrt(eating),data)
> summary(model)

Call:
lm(formula = log(expense) ~ income + sqrt(eating), data = data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.72270 -0.12522 -0.00519  0.13741  0.51204 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.984e+00  1.612e-01  43.332  < 2e-16 ***
income       9.653e-05  1.375e-05   7.020 1.13e-10 ***
sqrt(eating) 1.638e-02  4.117e-03   3.977 0.000116 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2221 on 129 degrees of freedom
Multiple R-squared:  0.3421,	Adjusted R-squared:  0.3319 
F-statistic: 33.53 on 2 and 129 DF,  p-value: 1.877e-12
```

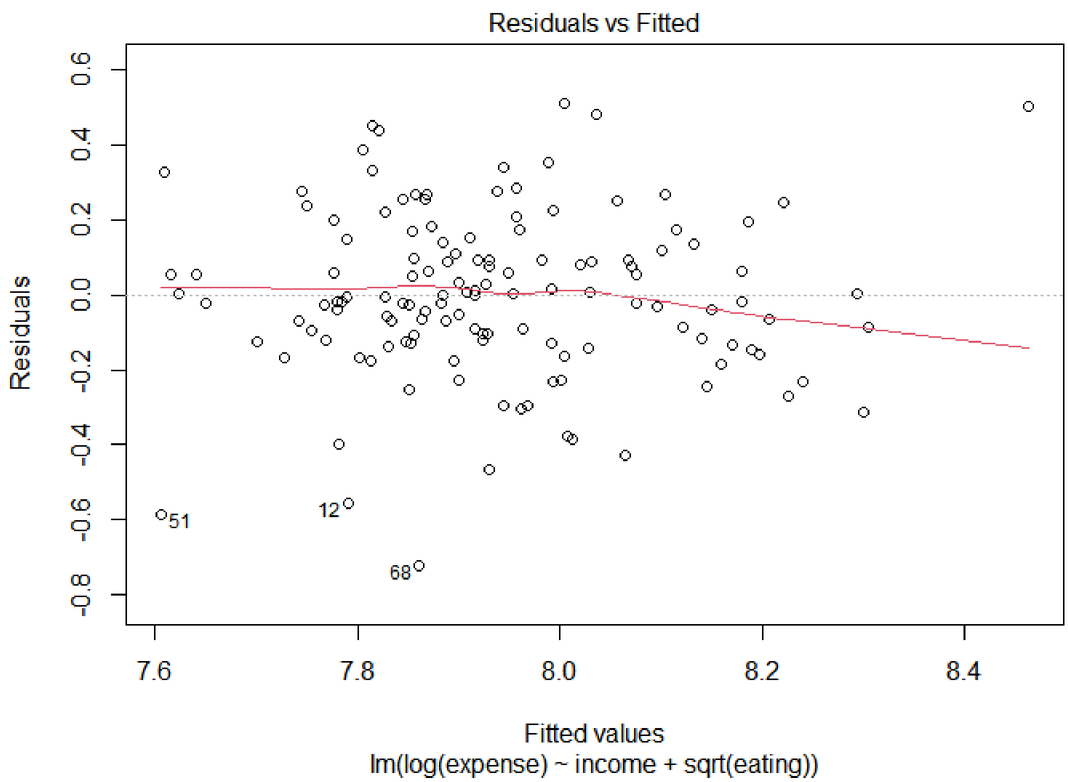
➔ Conclusion: Model after removing the DF Fit and DF Betas observations, we see: All coefficient standard errors in these models are little difference.

As a result, the influence points have an effect on estimated coefficient-Beta hat

c) Checking Outlier

```
> model=lm(log(expense) ~ income + sqrt(eating),data)
> plot(model)
Hit <Return> to see next plot: return
```

- Result



➔ Conclusion: According to this diagram, we see 3 point (12, 51, 68) which are:

	Expense	income	Eating
12	1390	2500	1195
51	1120	1500	849
68	1260	2500	1508

➔ Conclusion: These data have ‘expense’ smaller than others. We should not drop this observation because of losing the practicality of the model.

D. Summarize and answer the questions

1. What factors have the most impact on student spending?

Variable	Sup	income	home	eating	gen	scholarship
R^2	0.01326	0.289	0.001194	0.07664	0.01082	0.002095

a. Considering each variable affecting expense separately, the value of R^2 (R Square) reflects the degree of influence of the independent variables on the dependent variable. The variation of these two values is from 0 to 1. The closer to 1, the more meaningful the model is.

- In Table, we see that the R^2 value of the model has only the largest independent variable income (0.289) and the second largest is the independent variable eating (0.07664). Therefore, the factor that has the most impact on student spending is the part-time income of each student and then the cost of meals .
- For the monthly income, because the survey respondents are students, they often have different and unstable part-time jobs, most of them change jobs after about 2-3 months of overtime, or You can do many things in a certain amount of time. It is these things that make the income of students have a monthly difference. For food and drink, it has a big impact on spending because each month is different because you often have different parties or exchange meals with friends, so your monthly food costs are different.

b. What is the difference in spending between male and female students?

On average, male students spend 2994369 VND per month.

On average, female students spend 2821299 VND per month.

=> The difference in expenditure between male and female students is not much.

c) Does Home factor have a big impact on student spending? (Compare model with housing variable with model without housing variable -> compare based on residual standard error, adjusted R square)

```
> model<-lm(expense~.,data)
> op=summary(model)
> op$adj.r.squared
[1] 0.3442876
> ml4<-lm(expense~sup+income+eating+gen+scholarship,data)
> op4=summary(ml4)
> op4$adj.r.squared
[1] 0.3444164
```

Adjusted R Squared of model has home variable is 0.3442876.

Adjusted R Squared of model has not home variable is 0.3444164.

- ⇨ The home variable (independent variable) does not have much impact on
- ⇨ the dependent variable (expense).

To sum up, it shows that the spending on student housing does not have too much impact on the total monthly expenditure of each student. The reason can be understood that the housing expenditure of each student does not vary much between months, usually a certain amount because the room rent is often constant for a long time.

E. Weight of each member

Nguyen Thuy Linh	25%
Dao Thi Hong Nhung	25%
Bui Thi Mai Lương	25%
Nguyen Thi Hoai Linh	25%

F. Data source

https://docs.google.com/spreadsheets/d/1tsDn4GtT17tzHYINstyv2dRKU7LdmkWGSYOpe3v9l3s/edit?fbclid=IwAR3BHAOZh-O6_65TaD4fn3QUrjKQPOxPste_Pk5T_v0rerT3tayEpkizJnw#gid=0

1	expense	sup	income	home	gen	eating	scholarship
2	2150	3500	3500	1	0	1249	1
3	2320	1500	1500	0	1	1970	0
4	7850	1500	8000	1	1	1864	1
5	2350	1500	1500	0	1	1850	0
6	2850	2500	3500	1	1	1063	0
7	2600	2500	3000	1	0	1920	1
8	2360	2500	2500	1	0	1166	1
9	2910	1500	5000	1	0	1789	1
10	3010	5500	5500	1	0	1960	1
11	1612	2500	2500	1	0	1155	0
12	2070	1500	3000	1	1	1086	1
13	1390	2500	2500	1	1	1195	1
14	4020	4500	7000	1	1	1499	0
15	2350	1000	2000	0	0	1353	0
16	3610	2500	2500	1	0	1254	1
17	2375	3500	3500	1	1	958	1
18	2703	2500	5000	1	0	1714	0
19	2490	2500	2500	1	0	1635	0
20	2260	2500	2500	1	0	1445	1