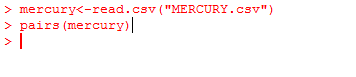
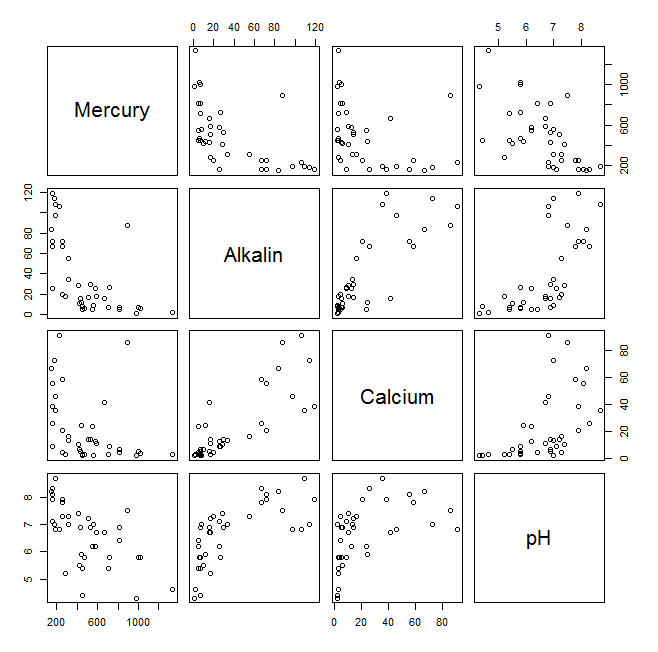
**Homework 3**

**STAT: 4510: 0001**

**Yubing Li 00808366**

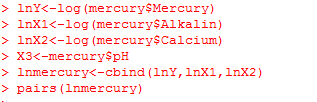
**1. Mercury Contamination**

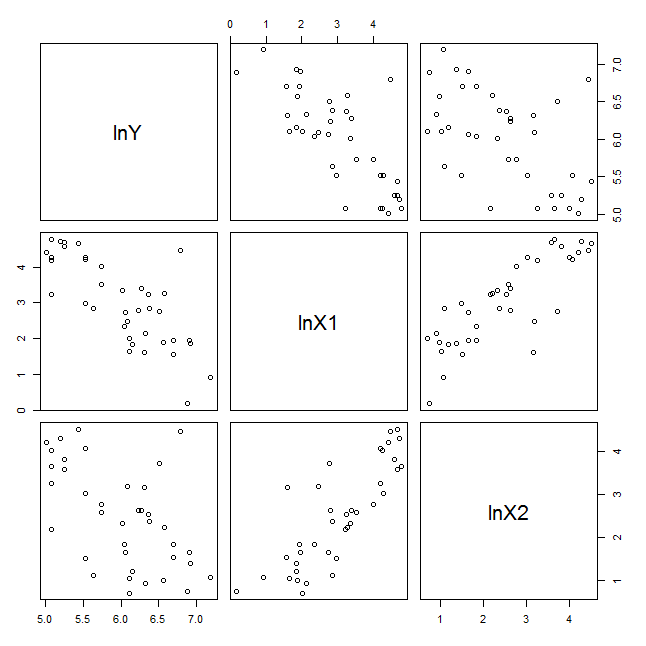
**(a)** **Scatterplot matrix:**



The response variable is Y: Mercury.

The explanatory variables are X1: Alkalin; X2: Calcium; X3: pH.

**(b) Scatterplot matrix of log:**

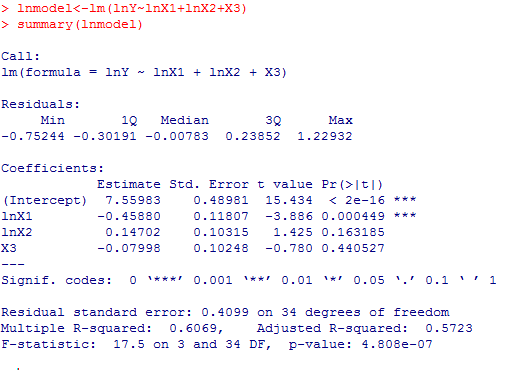


For lnY and lnX1:

The scatterplot with the response variable lnY and the explanatory variable lnX1 is the one in the middle of the 1st row. It shows a moderately strong, negative linear relationship.

For lnY and lnX2:

The scatterplot with the response variable lnY and the explanatory variable lnX2 is the one on the left of the 1st row. The graph does not indicate a linear relationship between them.

**(c) Model:**

The model is: l = 7.5598 – 0.4588×log(X1) + 0.1470×log(X2) -0.07998×X3

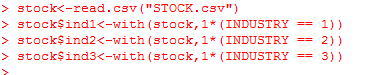
**(d) Interpret :**

log(Y) = 7.5598 – 0.4588×0 + 0.1470×0 – 0.07998×X3

Y = e7. 5598 – 0.4588×log(X1) + 0.1470×log(X2) ÷e0.07998×X3

β3 is the expected change in standardized mercury level related to per unit change in pH (X3), assuming all other explanatory variables are held fixed. is the unbiased estimator for β3.

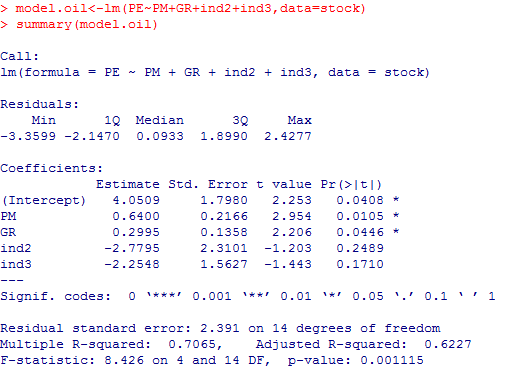
For per unit change in pH, the standardized mercury level is going to be divided by 1.0833.

**2. Stocks**

**(a) Linear model**

= 4.0509 + 0.6400(ProfitMargin) + 0.2995(GrowthRate) – 2.7795(ind2) – 2.2548(ind3)

with the oil industry being the base level.



**(b) Interpret the coefficient of indicator for ind3:**

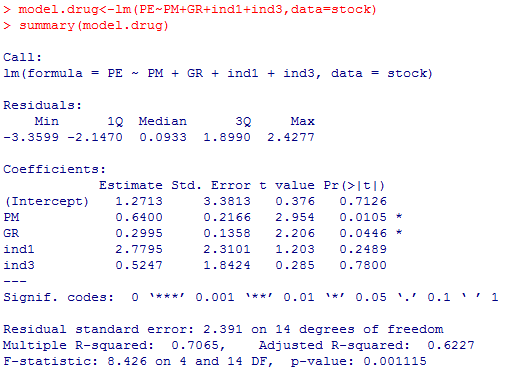
We can interpret the regression coefficient associated with ind3, -2.2548, to be the difference between the “computer/electronics” industry and the “oil” industry, holding other explanatory variables fixed.

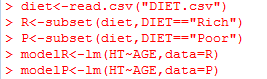
**(c) Drug/healthcare base level:**

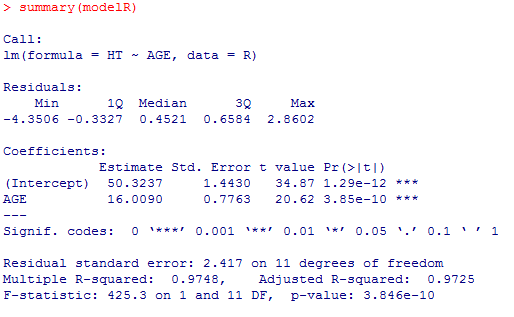
|  |  |  |
| --- | --- | --- |
| Oil | 4.0509 | 4.0509 |
| Drug/healthcare | 4.0509-2.7795 | 1.2714 |
| Computer/elec. | 4.0509-2.2548-1.2714 | 0.5247 |

= 1.2714 + 0.6400(ProfitMargin) + 0.2995(GrowthRate) + 2.7795(ind1) + 0.5247(ind3)

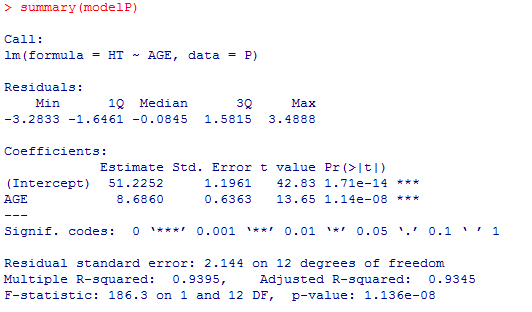
with the drug/healthcare industry being the base level

(Check the answer by fitting the new model, verified)

**3. Diet**

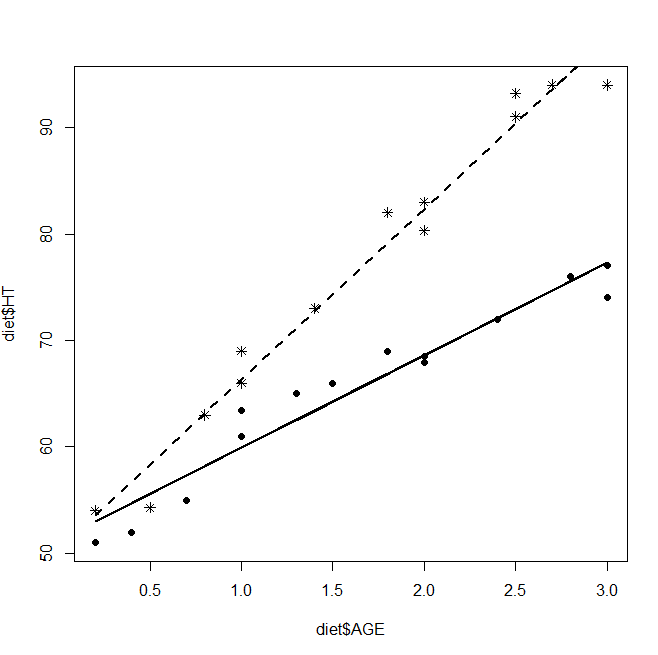
**(a) Separate LSRLs**

LSRL of HT on AGE for Protein-Rich: rich = 50.3237 + 16.0090 × (Age)



LSRL of HT on AGE for Protein-Poor: poor = 51.2252 + 8.6860 × (Age)

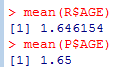
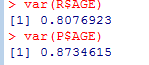
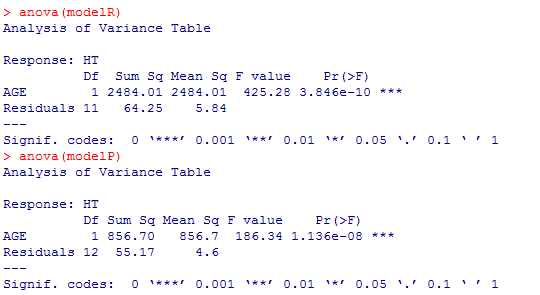
**(b) Single scatterplot**

 LSRL of HT on AGE for Protein-Rich

LSRL of HT on AGE for Protein-Poor

\* Group of Protein-Rich

● Group of Protein-Poor

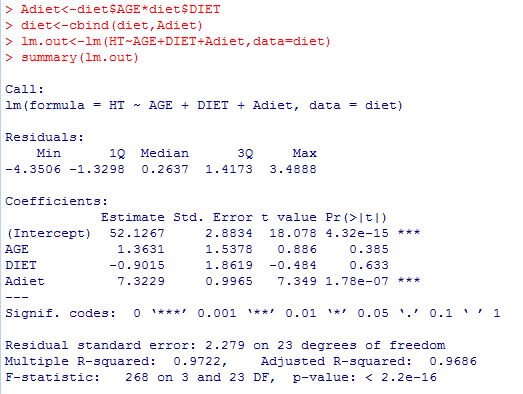
**(c) (d) (e) on the handwritten page.**

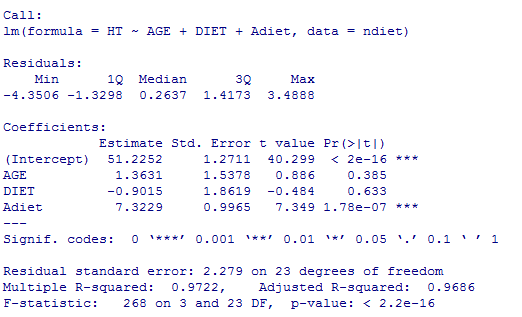
**4. Single regression model for Problem 3:**

**(a) Model is:**

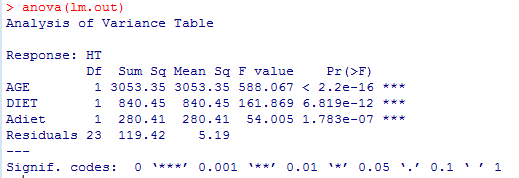
= β0 + β1×(Age) + β2×(Diet) + β3×( Age\* Diet) + ε

with indicator variable Diet = 1 if it’s Protein-Rich; Diet = 0 if it’s Protein-Poor.

**(b) Fit the model:**



= 51.2252+ 1.3631×(Age) – 0.9015×(Diet) + 7.3229×( Age\* Diet)

**(c) (d) (e) on the handwritten page.**