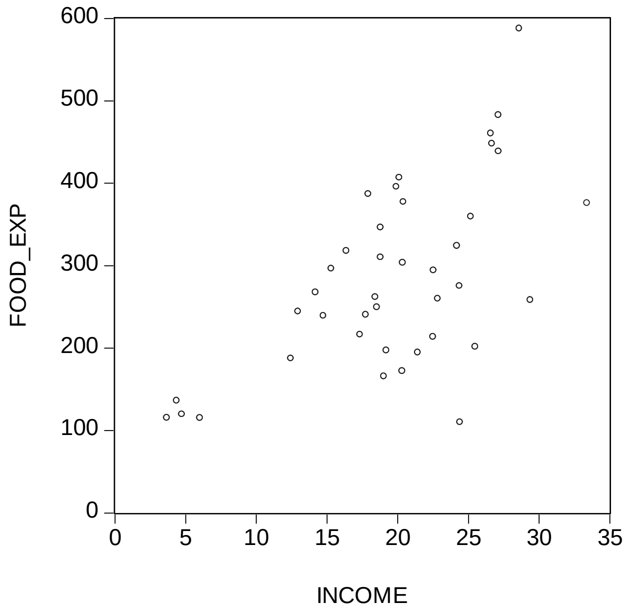
Introduction to Econometrics (Summer 2010) – Prof. Brauer and Prof. Dunne
EBA Programme | Faculty of Economics | Chulalongkorn University | Bangkok, Thailand

Sample annotation for assignments

Student name: ____________________
Student ID: ______________________
Chapter 02 – Simple linear regression



The problem is to estimate a linear regression for sample data for N=40 households for food expenditure (FOOD_EXP), given income (INCOME) data. The data are in food.wf1. A scatter plot of the data is in Figure 1. It appears that as INCOME increases, so does FOOD_EXP. This leads us to expect a positive regression slope coefficient for b2.

Note that the left axis has been rescaled to start at $0. (EViews' automatic plot starts at $100.) On the horizontal axis, INCOME is measured in hundreds of dollars, so that for example, $10 => $1,000.

Figure 1: Scatter plot of FOOD_EXP and INCOME

Some descriptive statistics (extract of copy/paste from Eviews) are displayed in Table 1. (The highlighted numbers are referred to later on.)

Table 1: Descriptive statistics

	INCOME	FOOD_EXP
Mean	19.60475	283.5735
Median	20.03000	264.4800
Maximum	33.40000	587.6600
Minimum	3.690000	109.7100
Std. Dev.	6.847773	112.6752
Sum	784.1900	11342.94
Sum Sq. Dev.	1828.788	495132.2
Observations	40	40

Average, or mean, INCOME in this sample of N=40 households is $19.60475 but since this is measured in hundreds of dollars, this means $1,960.48/month (rounded). Average FOOD_EXP = $283.57/week (rounded). The "sum sq. dev." for INCOME (1828.788) appears, for example, on p. 22 of the textbook in the calculation of the slope coefficient, b2.

The least squares regression equation is estimated as presented in Table 2. Items highlighted in yellow are mentioned below.

Table 2: Regression table

Dependent Variable: FOOD_EXP
Method: Least Squares
Date: 05/31/10 Time: 22:15
Sample: 1 40
Included observations: 40

	Coefficient	Std. Error	t-Statistic	Prob.
C	83.41600	43.41016	1.921578	0.0622
INCOME	10.20964	2.093264	4.877381	0.0000

R-squared	0.385002	Mean dependent var	283.5735
Adjusted R-squared	0.368818	S.D. dependent var	112.6752
S.E. of regression	89.51700	Akaike info criterion	11.87544
Sum squared resid	304505.2	Schwarz criterion	11.95988
Log likelihood	-235.5088	Hannan-Quinn criter.	11.90597
F-statistic	23.78884	Durbin-Watson stat	1.893880
Prob(F-statistic)	0.000019		

The estimated equation is FOOD_EXP = 83.42 + 10.21*INCOME. In the coefficient column of Table 2, the intercept, or constant = 83.42 (rounded; this is b1). The slope estimate, b2, is 10.21 (rounded). The interpretation is that if INCOME increases by one unit (e.g., from $10 to $11, that is, from $1,000 to $1,100 since INCOME is measured in hundreds of dollars), then, on average, FOOD_EXP increases by $10.21 ("increases," because the sign of b2 is positive).

To predict FOOD_EXP if INCOME=$20, we write FOOD_EXP = 83.42 + (10.21)*(20) = $287.62. In words, on average (given the sample data), a household with $2,000/month income would be expected to spend $287.62 on food/week.

Note that "Mean dependent var" and "S.D. dependent var" in Table 1 and Table 2 are the same (as they should be). There are two regression estimates, b1 and b2, therefore the degrees of freedom (df's) are 38 (N=40, minus 2). The error variance, sigma-hat squared, equals the sum of the squared residuals divided by the df's: 304,505.2 / 38 = 8,013.29 (p. 35 of the textbook). Taking the square root yields the "S.E. of regression" (standard error of regression) = 89.517, as seen in the regression output.

The variance-covariance matrix (food eq => view => covariance matrix) is:

	C	INCOME
C	1884.442	-85.90316
INCOME	-85.90316	4.381752

Thus, the variance of the intercept (or constant, or b1) = 1,884.442, the square root of which is the standard error, se(b1) = 43.410 (see regression output, Table 2). Likewise, the variance (b2) = 4.381752 => se(b2) = 2.093 (see Table 2). Dividing coefficient/se = t-statistic. For example, 10.20964/2.093264 = 4.87738 (rounded; see Table 2).

To receive a grade of outstanding (the full 3% marks for the assignment), you <u>MUST</u> to do something "extra." For example:

To show that the regression results do NOT change when the units of measurement change, we use the Eviews command line series income100 = income*100 to create a "new" income variable INCOME100 and then rerun the regression. That is, the INCOME values are multiplied times 100, so that INCOME=$10 now is INCOME100=$1,000. Table 3 shows the output.

Table 3: Regression output with INCOME100

Dependent Variable: FOOD_EXP
Method: Least Squares
Date: 06/02/10 Time: 16:05
Sample: 1 40
Included observations: 40

	Coefficient	Std. Error	t-Statistic	Prob.
C	83.41600	43.41016	1.921578	0.0622
INCOME100	0.102096	0.020933	4.877381	0.0000

R-squared	0.385002	Mean dependent var	283.5735
Adjusted R-squared	0.368818	S.D. dependent var	112.6752
S.E. of regression	89.51700	Akaike info criterion	11.87544
Sum squared resid	304505.2	Schwarz criterion	11.95988
Log likelihood	-235.5088	Hannan-Quinn criter.	11.90597
F-statistic	23.78884	Durbin-Watson stat	1.893880
Prob(F-statistic)	0.000019		

Comparing Table 2 with Table 3, it may be seen that the only numbers changed are those for b2 and se(b2). INCOME100 is interpreted as before: a <u>one-unit increase</u> in INCOME100 leads to a 0.102096 unit increase (note the positive sign) in FOOD_EXP. It follows that a <u>100-unit increase</u> in INCOME100 = 0.102096*100 = 10.2096 or about $10.21 which is the same result as obtained in Table 2. As regards prediction, note that FOOD_EXP = 83.42 + (0.102096)*(2000) = $287.61 (as before, except for a rounding error).

Thus, the results of the regression estimation are invariant with respect to the units of measurement used. This is further shown by comparing the regression residual plots for INCOME and INCOME100 below which are identical except for the income unit of measurement.