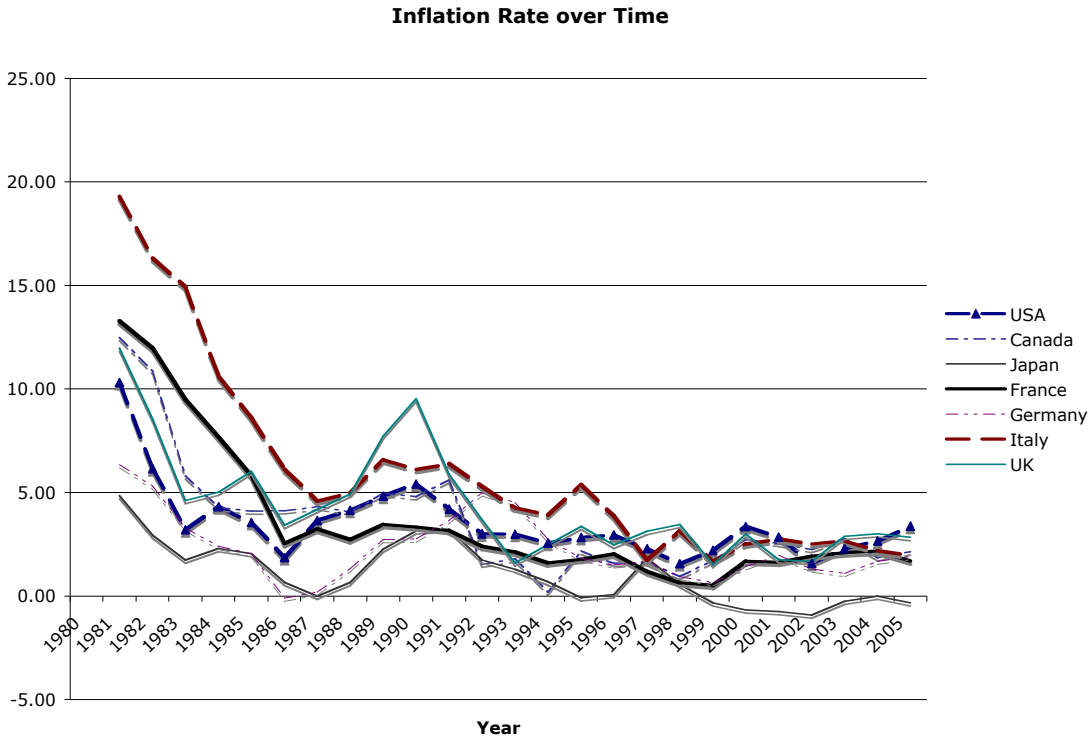
CHAPTER 1:
THE NATURE OF REGRESSION ANALYSIS

1.1 (a) These rates (%) are as follows. They are year-over-year, starting with 1981.

	USA	Canada	Japan	France	Germany	Italy	UK
1980							
1981	10.32	12.48	4.84	13.28	6.34	19.30	11.97
1982	6.16	10.86	2.94	11.97	5.31	16.31	8.53
1983	3.21	5.80	1.73	9.49	3.30	14.94	4.61
1984	4.32	4.28	2.30	7.67	2.39	10.62	5.01
1985	3.56	4.11	2.06	5.83	2.04	8.61	6.01
1986	1.86	4.13	0.67	2.53	-0.10	6.11	3.42
1987	3.65	4.32	0.00	3.24	0.19	4.59	4.18
1988	4.14	4.05	0.67	2.73	1.33	4.99	4.93
1989	4.82	4.95	2.27	3.46	2.73	6.59	7.72
1990	5.40	4.80	3.15	3.34	2.75	6.12	9.53
1991	4.21	5.61	3.23	3.16	3.65	6.39	5.87
1992	3.01	1.54	1.74	2.41	4.99	5.30	3.70
1993	2.99	1.79	1.28	2.14	4.50	4.25	1.60
1994	2.56	0.20	0.68	1.60	2.74	3.92	2.48
1995	2.83	2.16	-0.08	1.78	1.83	5.37	3.36
1996	2.95	1.59	0.08	2.02	1.50	3.87	2.46
1997	2.29	1.63	1.84	1.19	1.70	1.75	3.12
1998	1.56	0.96	0.58	0.65	0.94	3.15	3.46
1999	2.21	1.71	-0.33	0.52	0.65	1.66	1.52
2000	3.36	2.74	-0.66	1.68	1.43	2.52	2.99
2001	2.85	2.55	-0.74	1.65	1.97	2.76	1.75
2002	1.58	2.25	-0.92	1.94	1.31	2.52	1.67
2003	2.28	2.78	-0.25	2.08	1.09	2.66	2.90
2004	2.66	1.86	0.00	2.16	1.69	2.19	3.00
2005	3.39	2.15	-0.34	1.70	1.92	1.95	2.83

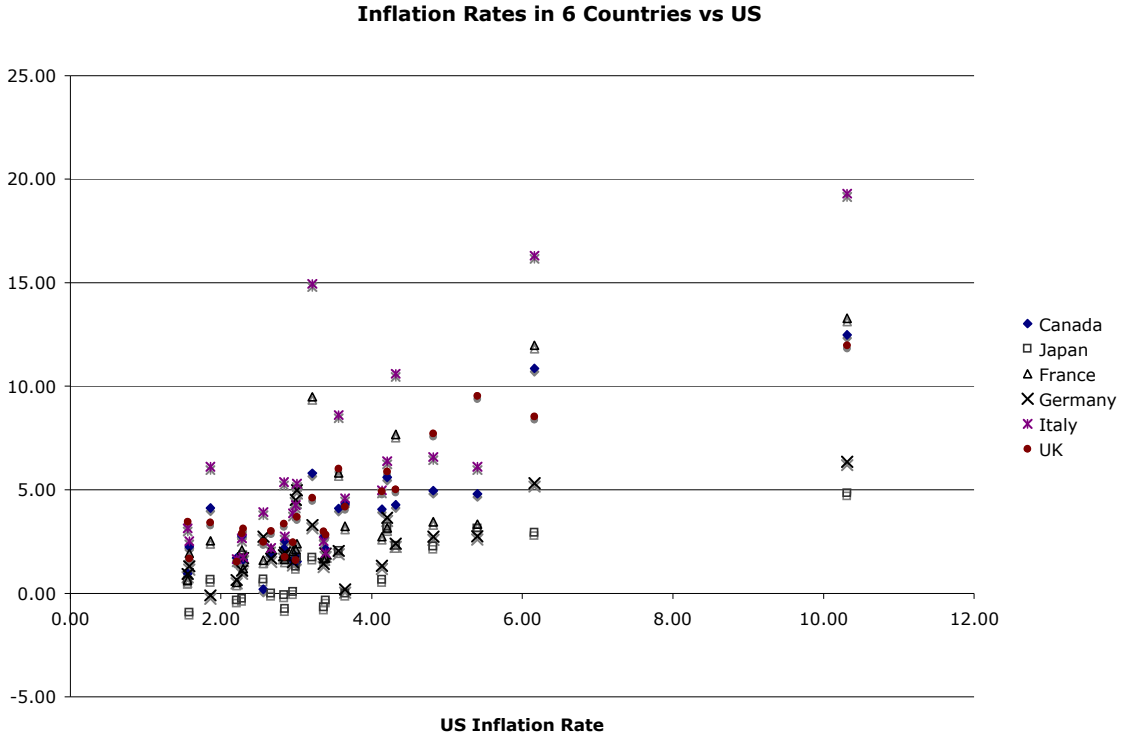
(b)



(c) As you can see from this figure, the inflation rate of each of the countries has *generally* declined over the years.

(d) As a measure of variability, we can use the standard deviation. These standard deviations are 1.81, 2.85, 1.49, 3.40, 1.60, 4.70, and 2.65, respectively, for the US, Canada, Japan, France, Germany, Italy, and the UK. The highest variability is thus found for Italy and the lowest for Japan.

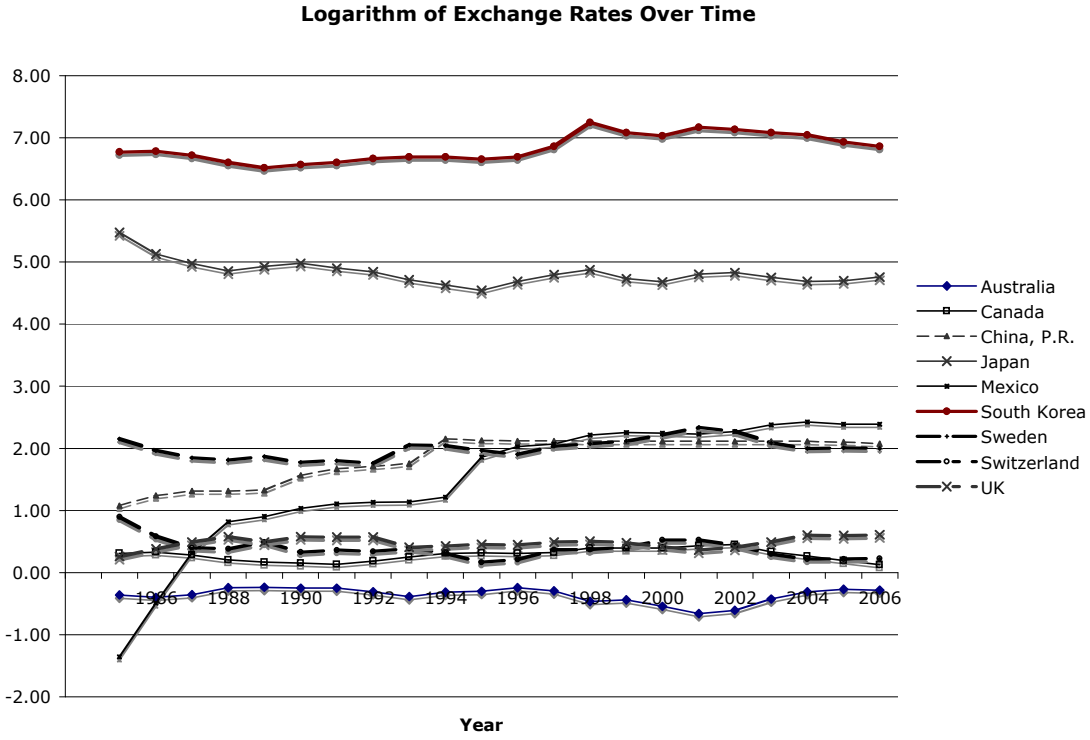
1.2. (a) The graph of the inflation rates of the six countries plotted against the US inflation rate is as follows:



(b) As the figure shows, in general the inflation rates of the six countries are positively correlated with the US inflation rate.

(c) Remember that correlation does not mean causation. One may have to consult a book on international macroeconomics to find out if there is any causal connection between the US and the other countries' inflation rates.

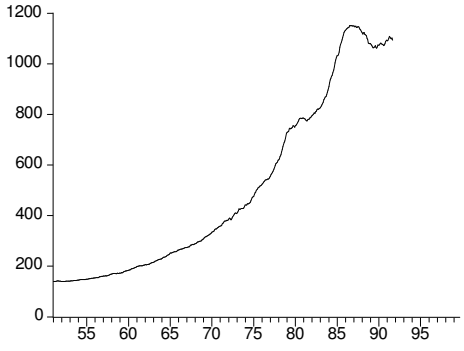
1.3 (a) For better visual impression the logarithm of the exchange rate is plotted on the vertical axis and time on the horizontal axis.



As you can see, the exchange rates show a good deal of variability. For example, in 1985 one US dollar only bought about 0.257 Pesos, but in 2004 it could buy about 11.29 Pesos.

(b) Again, the picture is mixed. For instance, between 1985 and 2006, the U.S. dollar appreciated at a relatively high rate against the Peso, but for most of the other currencies the relationship more slowly and steadily increased.

1.4. The graph of the M1 money supply is as follows:



M1 Money Supply: USA, 1951:01-1999:09

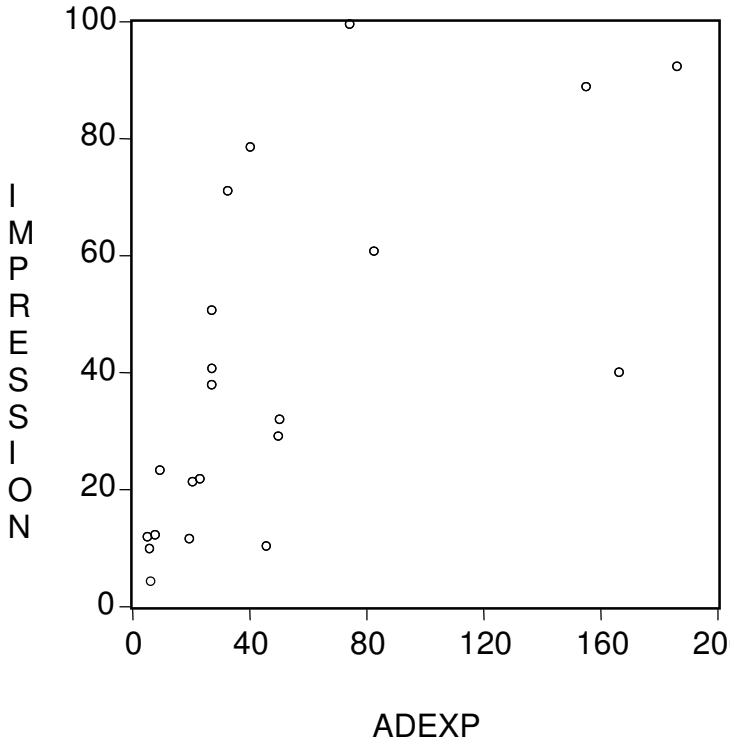
As GDP increases over time, naturally a higher amount of the money supply is needed to finance the increased output.

1.5. Some of the relevant variables would include: (1) wages or earnings

in criminal activity, (2) hourly wages or earnings in non-criminal activity, (3) probability of getting caught, (4) probability of conviction, (5) expected sentence after conviction. Note that it may not be easy to get data on earnings in the illegal activities. Anyway, refer to the Becker article cited in the text.

1.6. One key factor in the analysis would be the labor force participation rate of people in the 65-69 age category. Data on labor force participation are collected by the Labor Department. If, after the new law went into effect, we find increased participation of these "senior" citizens in the labor force, that would be a strong indication that the earlier law had artificially restricted their labor market participation. It would also be interesting to find out what kinds of of jobs these workers get and what they earn.

1.7 (*a*), (*b*) & (*c*). As the following figure shows, there seems to be a positive relationship between the two variables, although it does not seem to be very strong. This probably suggests that it pays to advertise; otherwise, it is bad news for the advertising industry.



CHAPTER 2:
TWO-VARIABLE REGRESSION ANALYSIS: SOME BASIC IDEAS

2.1 It tells how the mean or average response of the sub-populations of Y varies with the fixed values of the explanatory variable (s).

2.2 The distinction between the sample regression function and the population regression function is important, for the former is is an estimator of the latter; in most situations we have a sample of observations from a given population and we try to learn something about the population from the given sample.

2.3 A regression model can never be a completely accurate description of reality. Therefore, there is bound to be some difference between the actual values of the regressand and its values estimated from the chosen model. This difference is simply the stochastic error term, whose various forms are discussed in the chapter. The residual is the sample counterpart of the stochastic error term.

2.4 Although we can certainly use the mean value, standard deviation and other summary measures to describe the behavior the of the regressand, we are often interested in finding out if there are any causal forces that affect the regressand. If so, we will be able to better predict the mean value of the regressand. Also, remember that econometric models are often developed to test one or more economic theories.

2.5 A model that is linear in the parameters; it may or may not be linear in the variables.

2.6 Models (a), (b), (c) and (e) are linear (in the parameter) regression models. If we let $\alpha = \ln \beta_1$, then model (d) is also linear.

2.7 (a) Taking the natural log, we find that $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$, which becomes a linear regression model.
(b) The following transformation, known as the **logit** transformation, makes this model a linear regression model:

$$\ln [(1- Y_i)/Y_i] = \beta_1 + \beta_2 X_i + u_i$$

(c) A linear regression model
(d) A nonlinear regression model
(e) A nonlinear regression model, as β_2 is raised to the third power.

2.8 A model that can be made linear in the parameters is called an intrinsically linear regression model, as model (a) above. If β_2 is 0.8 in model (d) of Question 2.7, it becomes a linear regression

model, as $e^{-0.8(X_i - 2)}$ can be easily computed.

2.9 (a) Transforming the model as $(1/Y_i) = \beta_1 + \beta_2 X_i$ makes it a linear regression model.
(b) Writing the model as $(X_i/Y_i) = \beta_1 + \beta_2 X_i$ makes it a linear regression model.
(c) The transformation $\ln[(1 - Y_i)/Y_i] = -\beta_1 - \beta_2 X_i$ makes it a linear regression model.
Note: Thus the original models are intrinsically linear models.

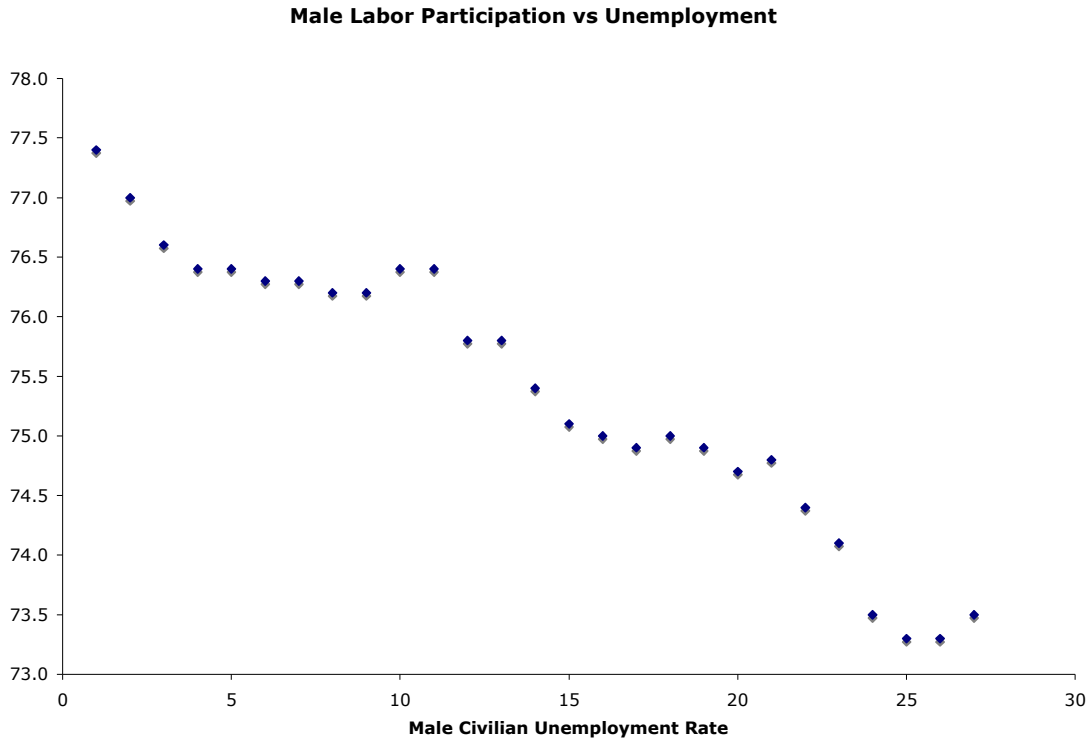
2.10 This scattergram shows that more export-oriented countries on average have more growth in real wages than less export oriented countries. That is why many developing countries have followed an export-led growth policy. The regression line sketched in the diagram is a sample regression line, as it is based on a sample of 50 developing countries.

2.11 According to the well-known Heckscher-Ohlin model of trade, countries tend to export goods whose production makes intensive use of their more abundant factors of production. In other words, this model emphasizes the relation between factor endowments and comparative advantage.

2.12 This figure shows that the higher is the minimum wage, the lower is per head GNP, thus suggesting that minimum wage laws may not be good for developing countries. But this topic is controversial. The effect of minimum wages may depend on their effect on employment, the nature of the industry where it is imposed, and how strongly the government enforces it.

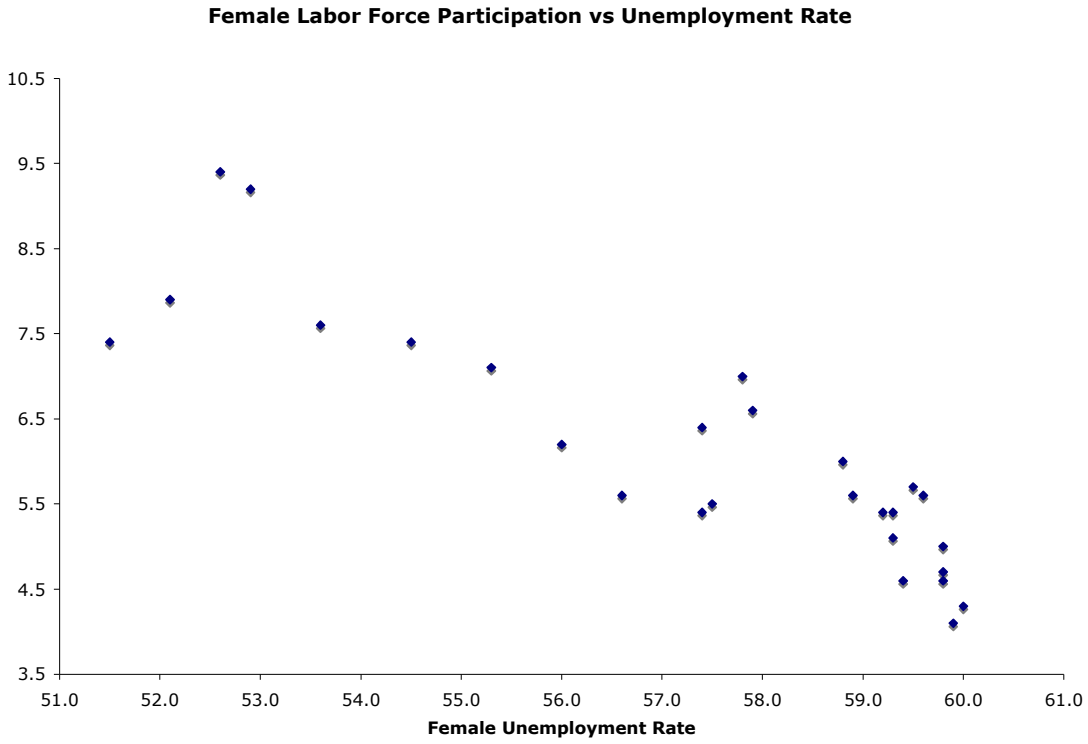
2.13 It is a sample regression line because it is based on a sample of 15 years of observations. The scatter points around the regression line are the actual data points. The difference between the actual consumption expenditure and that estimated from the regression line represents the (sample) residual. Besides GDP, factors such as wealth, interest rate, etc. might also affect consumption expenditure.

2.14 (a) The scattergram is as follows:



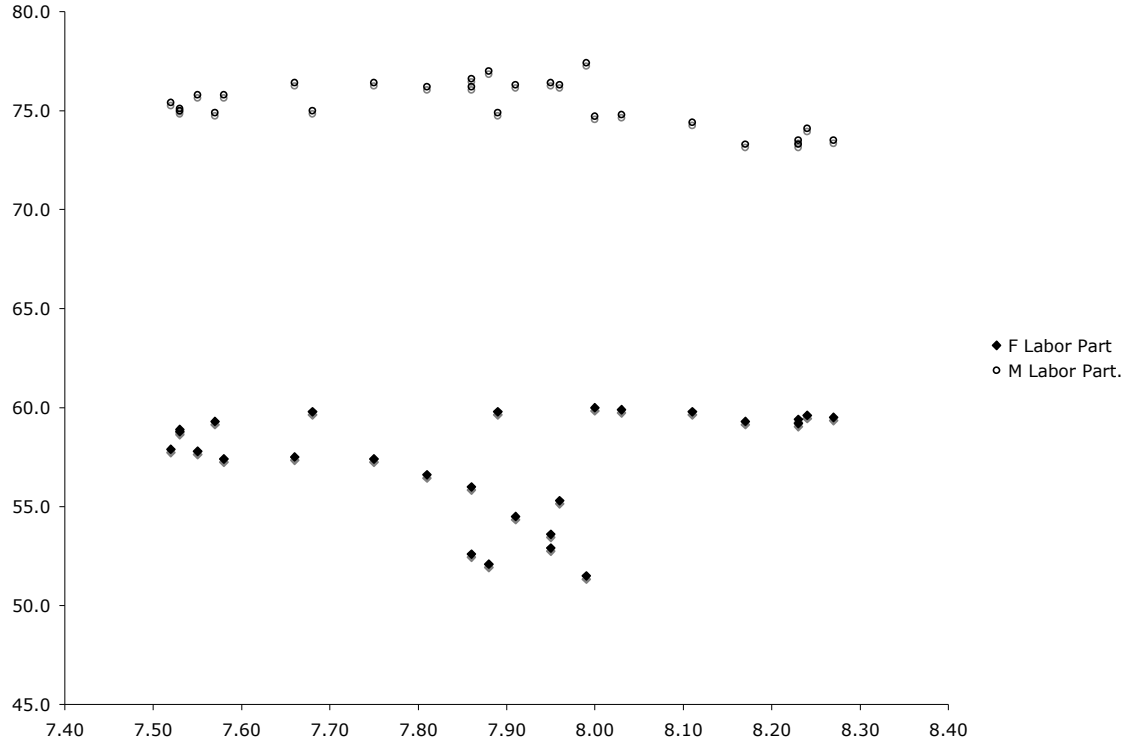
The negative relationship between the two variables seems seems relatively reasonable. As the unemployment rate increases, the labor force participation rate decreases, although there are several minor peaks and valleys in the graph.

(b) The scattergram is as follows:



Here the **discouraged worker hypothesis** of labor economics seems to be at work: unemployment discourages female workers from participating in the labor force because they fear that there are no job opportunities.

(c) The plot of Male and Female Labor Force Participation against AH82 shows the following:



There is a similar relationship between the two variables for males and females, although the Male Labor Participation Rate is always significantly higher than that of the Females. Also, there is quite a bit more variability among the Female Rates. To validate these statements, the average Male Rate is 75.4 %, whereas the Female average is only 57.3 %. With respect to variability, the Male Rate standard deviation is only 1.17 %, but the Female standard deviation is 2.73 %, more than double that of the Male Rate. Keep in mind that we are doing simple bivariate regressions here. When we study multiple regression analysis, we may have some different conclusions.