Econ 301: Econometrics

Alpay Filiztekin

Sabanci University

Basic Data Handling

The structure of economic data

- Cross-sectional data (y_i, $i = 1, ..., N$)
- Time series data (y_t, $t = 1, ..., T$)
- Panel data (y_{it}, $i = 1, ..., N$ and $t = 1, ..., T$)

The structure of economic data

- Cross-sectional data (y_i, $i = 1, ..., N$)
- Time series data (y_t, $t = 1, ..., T$)
- Panel data (y_{it}, $i = 1, ..., N$ and $t = 1, ..., T$)

The structure of economic data

- Cross-sectional data (y_i, $i = 1, ..., N$)
- Time series data (y_t, $t = 1, ..., T$)
- Panel data (y_{it}, $i = 1, ..., N$ and $t = 1, ..., T$)

Looking at the raw data

First things first: look at the data!

Data Browser

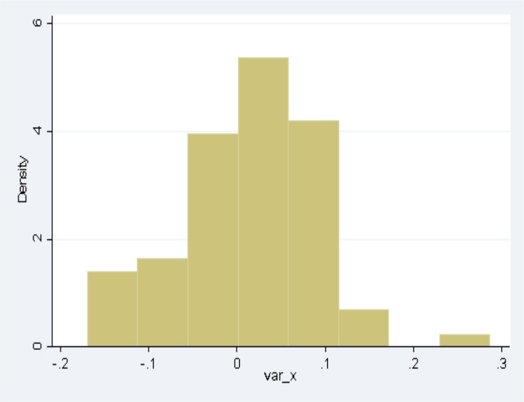
Preserve | Restore | Sort | << | >> | Hide | Delete...

my [58] = 561

	my	vestl	sise	sahol	ptofs	petkm	mutlu	vakko	tuprs	thyao	tcell	migrs	kent	kchol	isgyo	g
58	2006m10	3.614737	2.335789	5.436316	3.859474	5.268948	1.424211	.9357895	18.94421	6.33421	7.25421	15.41895	49.65211	3.369474	1.906316	64
59	2006m11	3.864545	2.311364	5.316818	3.957727	5.365909	1.465	1.022727	19.69455	6.370455	6.686818	14.78909	43.85318	3.524091	1.970909	61
60	2006m12	3.682857	2.229524	5.501429	3.891905	5.171429	1.570476	1.096667	19.74	6.190476	6.34619	15.63809	44.37905	3.461905	2.044762	62
61	2007m1	3.518	2.2065	5.4495	3.436	5.219	1.593	1.4235	19.838	6.8425	6.8205	16.8265	35.6165	3.888	2.0085	7
62	2007m2	3.675	2.518	5.783	3.8545	6.255	1.6285	1.3595	22.085	7.805	7.0905	17.4325	36.537	4.154	2.1795	
63	2007m3	3.386364	2.286818	5.394091	4.330909	6.445455	1.532273	1.219091	22.48955	7.156818	6.715	16.41455	36.84818	3.922727	2.056364	74
64	2007m4	3.646	2.336	5.9675	4.518	7.3275	1.864	1.228	24.4045	8.0125	7.128	16.852	39.314	4.505	2.2285	7
65	2007m5	3.553043	2.260435	5.925218	4.967391	8.436956	1.890435	1.146522	25.97565	8.51087	7.749565	16.49	45.54826	4.553913	2.132174	87
66	2007m6	3.185714	2.345714	6.097619	4.876667	8.866667	2.308571	1.417143	26.35952	8.597619	8.12381	19.09714	48.97952	4.608572	2.003809	87
67	2007m7	3.266364	2.641818	6.922273	5.025909	9.590476	2.464091	1.458636	28.65	9.904546	8.768182	20.79	55.65091	5.135	2.112727	86
68	2007m8	2.983636	2.389545	6.617273	5.142273	9.370455	2.362727	1.265	25.86091	8.895454	8.710455	19.41773	53.61136	4.791818	1.918182	
69	2007m9	3.118	2.596	7.1425	5.156	9.9125	2.765	1.173	27.272	8.9775	9.34	20.3545	54.9835	5.3975	1.9495	
70	2007m10	3.483809	2.577	7.588095	5.519524	9.635715	2.625714	1.164286	27.56429	9.007143	10.45	19.67714	49.57476	5.507619	1.964286	87
71	2007m11	3.25	.	6.82	5.24	9.363636	2.488182	1.103182	27.6	8.134091	11.63727	20.34546	45.90409	5.43	1.793636	89
72	2007m12	3.034737	.	6.446842	5.400526	9.55	2.463684	1.071053	28.08368	8.652632	12.76158	21.58105	44.26579	.	1.686842	91
73	2008m1	2.407273	.	5.670909	4.706364	7.681818	2.380455	.8877273	27.91	7.784091	11.13591	19.91909	37.82318	.	1.516364	93
74	2008m2	2.004762	.	5.406667	4.76	6.704762	2.095714	.802381	26.09762	7.535714	10.93095	18.74476	37.46333	.	1.339048	10
75	2008m3	1.861429	.	4.664762	5.105238	7.359524	1.983809	.7833333	25.96143	6.397619	11.15238	19.08714	37.23714	.	1.08	1
76	2008m4	1.899091	.	4.269091	4.951818	6.872727	1.901818	.9690909	27.25	6.818182	11.12545	19.03727	38.90091	.	1.028182	10

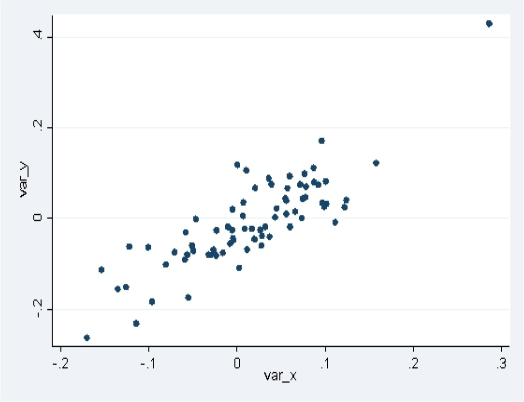
Graphical Analysis

- ***Histograms*** give you an impression of the distribution of a variable.
  In Stata type: `histogram varx`



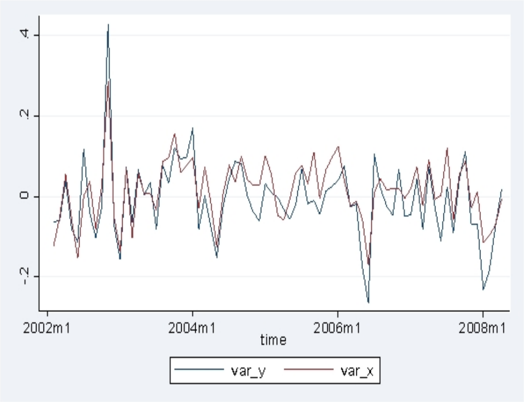
Graphical Analysis

- ***Scatter plots*** give combination of values from two series for the purpose of determining their relationship.
  In Stata type: `scatter varx vary`



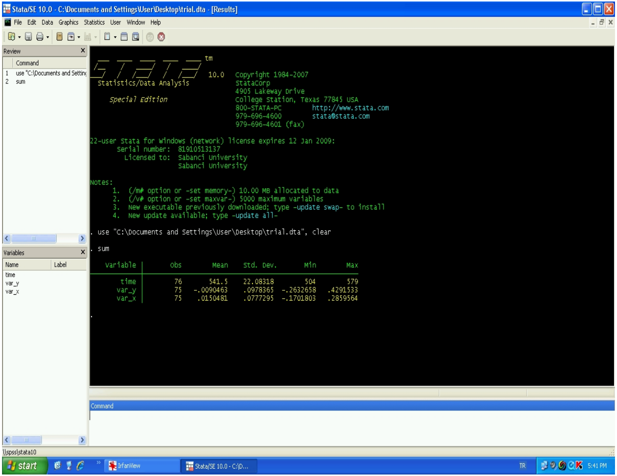
Graphical Analysis

- ***Line grpahs*** facilitate comparison of series.
  In Stata type: `line varx vary time`



Summary Statistics

To gain more insight about the distribution of variables.



Data Transformation: Real vs. Nominal

Nominal series incorporate a price component that can obscure the fundamental features of the series we are interested in.

To circumvent the problem, one can convert nominal series into real terms by using an appropriate price deflator (e.g. dividing it by consumer price index).

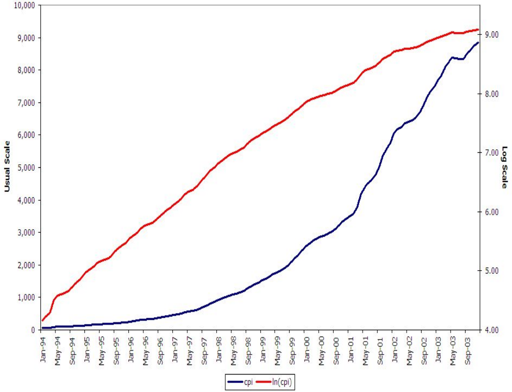
Data Transformation: Real vs. Nominal

Nominal series incorporate a price component that can obscure the fundamental features of the series we are interested in.

To circumvent the problem, one can convert nominal series into real terms by using an appropriate price deflator (e.g. dividing it by consumer price index).

Data Transformation: Logs

Taking logarithms linearizes:



Growth Rates and Differencing

Let the growth rate of (percentage change in) a variable z between time periods t and $t+1$ be g:

$$z_{t+1} = (1+g)z_t \tag{1}$$

Then

$$g = \frac{z_{t+1} - z_t}{z_t} = \frac{z_{t+1}}{z_t} - 1 \tag{2}$$

Growth Rates and Differencing

The logarithmic approximation of g will be:

$$g \approx ln(\frac{z_{t+1}}{z_t}) = ln(z_{t+1}) - lnz_t \equiv \triangle ln(z_{t+1}) \tag{3}$$

Growth Rates and Differencing

Notice that average growth rate of a variable *z* between time periods *t* and *t+s* is:

$$z_{t+s} = (1+g)^s z_t \tag{4}$$

Then average growth rate, *g* is:

$$g = (\frac{z_{t+1}}{z_t})^{(1/s)} - 1 \tag{5}$$

Growth Rates and Differencing

Again logarithmic approximation will be:

$$g \approx \frac{1}{s} * ln(\frac{z_{t+1}}{z_t}) = \frac{1}{s} * (ln(z_{t+1}) - lnz_t) = \frac{1}{s} * \triangle(ln(z_{t+1})) \quad (6)$$

What Is A Regression?

Regression and Correlation

Regression is not correlation.

Although it implies correlation.

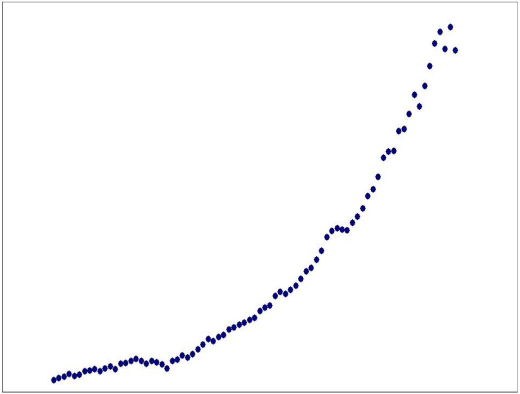
Regression and Correlation

Regression is not correlation.

Although it implies correlation.

Regression and Correlation

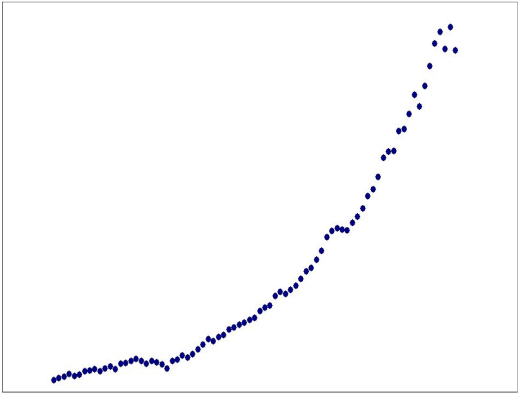
- Are these two variables correlated?



- Yes, the correlation coefficient is 0.9308.

Regression and Correlation

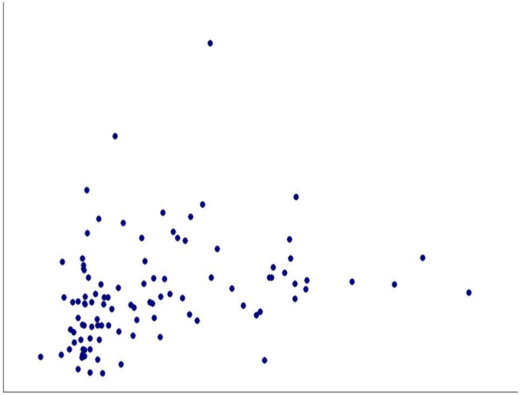
- Are these two variables correlated?



- Yes, the correlation coefficient is 0.9308.

Regression and Correlation

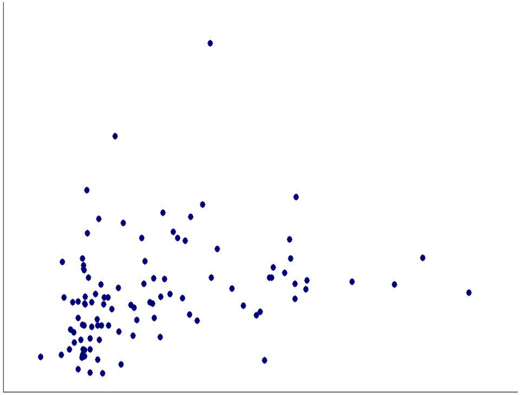
- Are these two variables correlated?



- Somewhat, the correlation coefficient is 0.3048.

Regression and Correlation

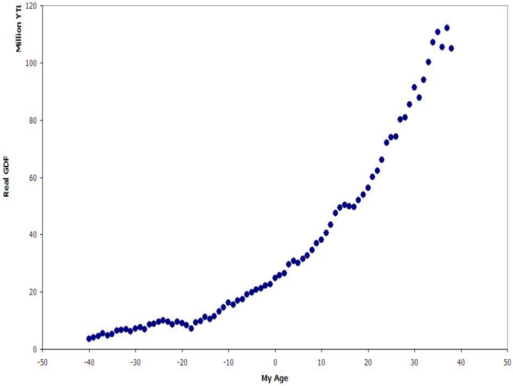
- Are these two variables correlated?



- Somewhat, the correlation coefficient is 0.3048.

Regression and Causation

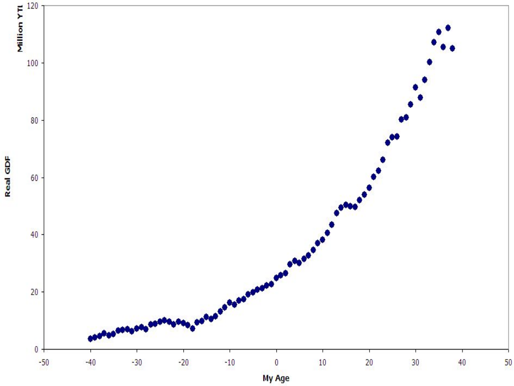
- Are these two variables causally related?



- I don't think so!!.

Regression and Causation

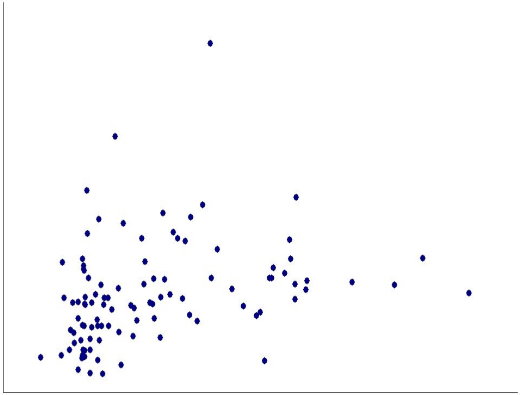
- Are these two variables causally related?



- I don't think so!!.

Regression and Correlation

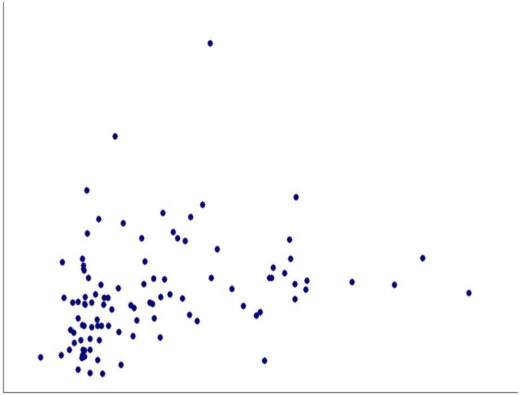
- Are these two variables causally related?



- Who knows?

Regression and Correlation

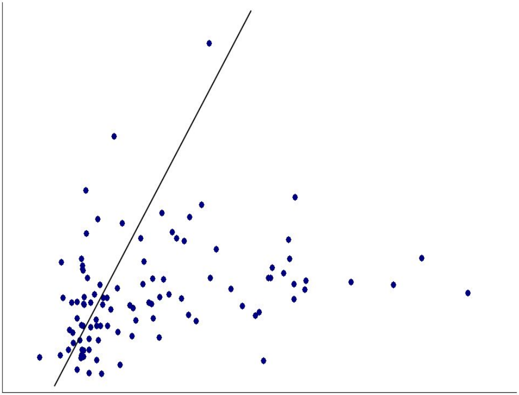
- Are these two variables causally related?



- Who knows?

Regression and Correlation

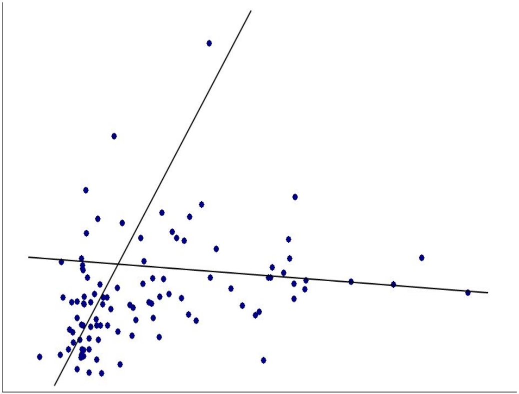
Are these two variables causally related?



Who knows?

Regression and Correlation

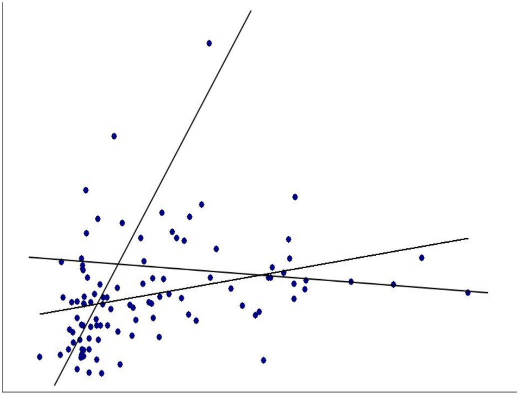
Are these two variables causally related?



Who knows?

Regression and Correlation

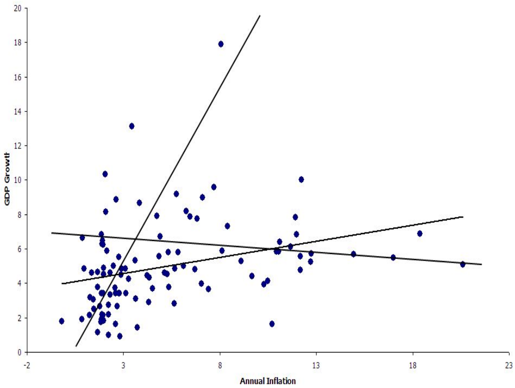
Are these two variables causally related?



Who knows?

Regression and Correlation

Are these two variables causally related?



Who knows?

What is Regression?

The meaning in the dictionary is:

1. **Reversion to earlier state:** a return to an earlier or less developed condition or way of behaving.
2. **Movement backward:** a going backward or a backward movement or progress, especially through the earlier stages or forms of something.

What is Regression?

The meaning in the dictionary is:

1. **Reversion to earlier state:** a return to an earlier or less developed condition or way of behaving.
2. **Movement backward:** a going backward or a backward movement or progress, especially through the earlier stages or forms of something.

What is Regression?

The meaning in the dictionary is:

1. **Reversion to earlier state:** a return to an earlier or less developed condition or way of behaving.
2. **Movement backward:** a going backward or a backward movement or progress, especially through the earlier stages or forms of something.

What is Regression?

The meaning in the dictionary is:

3. **Association between variables:** a process for determining the statistical relationship between a random variable and one or more independent variables that is used to predict the value of the random variable.

What is Regression?

History of regression analysis:

- The earliest form of **"regression"** was the method of ***least squares***, which was published by Legendre in 1805, and by Gauss in 1809. The term least squares is from Legendre's, ***"moindres carrés"***. However, Gauss claimed that he had known the method since 1795.
- Legendre and Gauss both applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the Sun. Euler had worked on the same problem without success. Gauss published a further development of the theoryof least squares in 1821,including a version of the ***Gauss-Markov theorem***.

What is Regression?

History of regression analysis:

- The term **regression** was coined in the nineteenth century to describe a biological phenomenon, namely that the progeny of exceptional individuals tend on average to be less exceptional than their parents and more like their more distant ancestors. Francis Galton studied this phenomenon and applied slightly misleading term ***"regression towards mediocrity"*** to it.
- For Galton, regression had only this biological meaning, but his work was later extended by Udny Yule and Karl Pearson to a more general statistical context.
- Nowadays the term **regression** is often synonymous with **least squares curve fitting**.