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 \((y_{it}, i = 1, ..., N \text{ and } t = 1, ..., T)\)
The structure of economic data

- Cross-sectional data \((y_i, \ i = 1, \ldots, N)\)
- Time series data \((y_t, \ t = 1, \ldots, T)\)
- Panel data \((y_{it}, \ i = 1, \ldots, N \text{ and } t = 1, \ldots, T)\)
The structure of economic data

- Cross-sectional data \((y_i, i = 1, \ldots, N)\)
- Time series data \((y_t, t = 1, \ldots, T)\)
- Panel data \((y_{it}, i = 1, \ldots, N \text{ and } t = 1, \ldots, T)\)
Looking at the raw data

First things first: look at the data!
Histograms give you an impression of the distribution of a variable.
In Stata type: `histogram varx`
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 graphs* facilitate comparison of series.

In Stata type: `line varx vary time`
Summary Statistics

To gain more insight about the distribution of variables.
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).
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).
Taking logarithms linearizes:
Let the growth rate of (percentage change in) a variable $z$ between time periods $t$ and $t+1$ be $g$:

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

Then

$$g = \frac{z_{t+1} - z_t}{z_t} = \frac{z_{t+1}}{z_t} - 1 \quad (2)$$
The logarithmic approximation of $g$ will be:

$$g \approx \ln\left(\frac{z_{t+1}}{z_t}\right) = \ln(z_{t+1}) - \ln z_t \equiv \Delta \ln(z_{t+1})$$ (3)
Notice that average growth rate of a variable $z$ between time periods $t$ and $t+s$ is:

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

Then average growth rate, $g$ is:

$$g = \left( \frac{Z_{t+1}}{Z_t} \right)^{(1/s)} - 1$$  \hspace{1cm} (5)
Again logarithmic approximation will be:

\[ g \approx \frac{1}{s} \cdot \ln \left( \frac{z_{t+1}}{z_t} \right) = \frac{1}{s} \cdot (\ln(z_{t+1}) - \ln z_t) = \frac{1}{s} \cdot \triangle(\ln(z_{t+1})) \] (6)
What Is A Regression?
Regression is not correlation.

Although it implies correlation.
Regression and Correlation

Regression is not correlation.

Although it implies correlation.
Are these two variables correlated?

Yes, the correlation coefficient is 0.9308.
Are these two variables correlated?

Yes, the correlation coefficient is 0.9308.
Are these two variables correlated?

Somewhat, the correlation coefficient is 0.3048.
Are these two variables correlated?

Somewhat, the correlation coefficient is 0.3048.
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!!.
Are these two variables causally related?

Who knows?
Are these two variables causally related?

Who knows?
Regression and Correlation

Are these two variables causally related?

Who knows?
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:

- **Reversion to earlier state**: a return to an earlier or less developed condition or way of behaving.
- **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 theory of 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.