Welcome!

Time Series Analytics

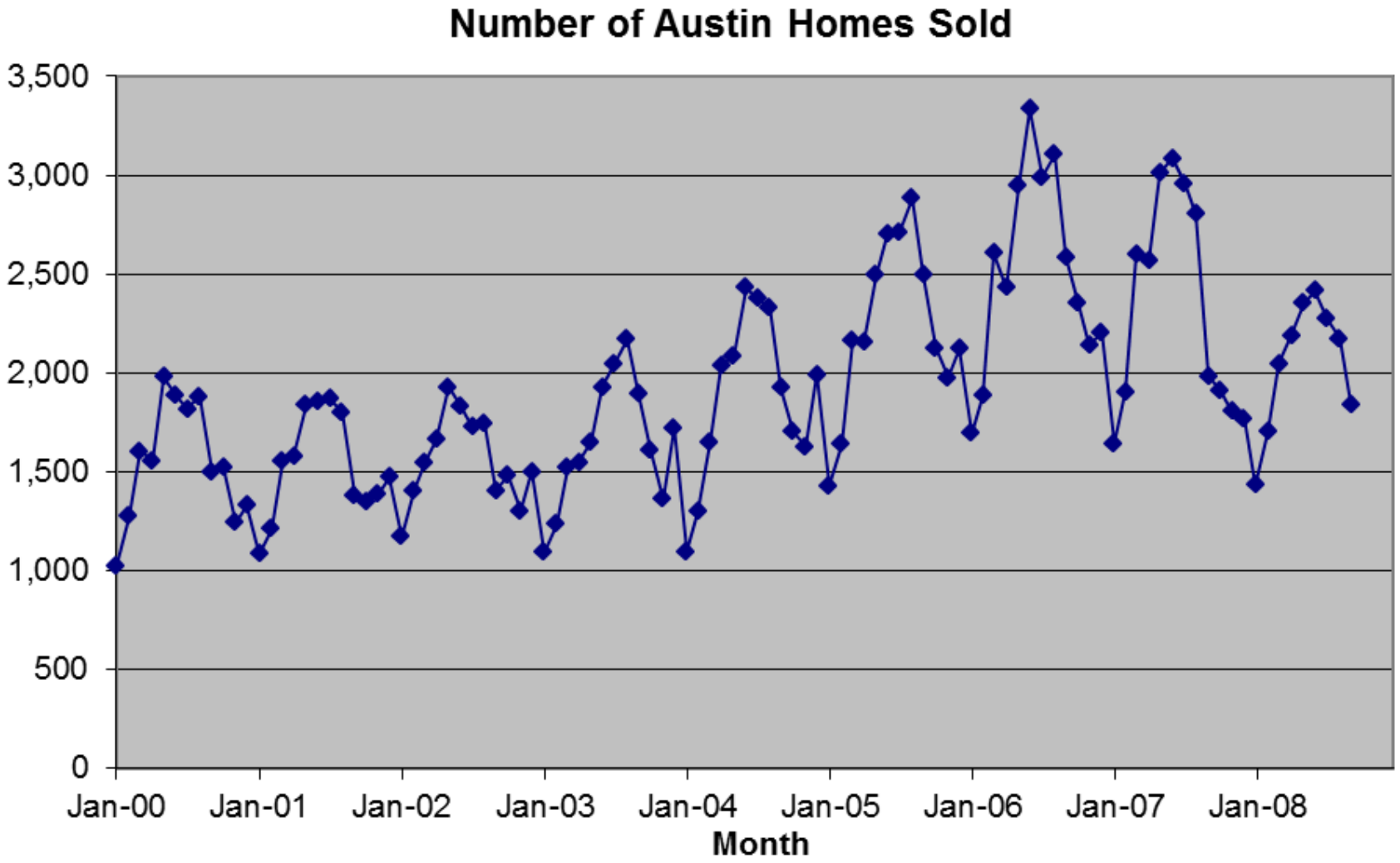
Professor Tom Sager

TomSager@mail.utexas.edu

Time Series

Definition. A time series is a sequence of numbers in chronological order.

A Real Time Series



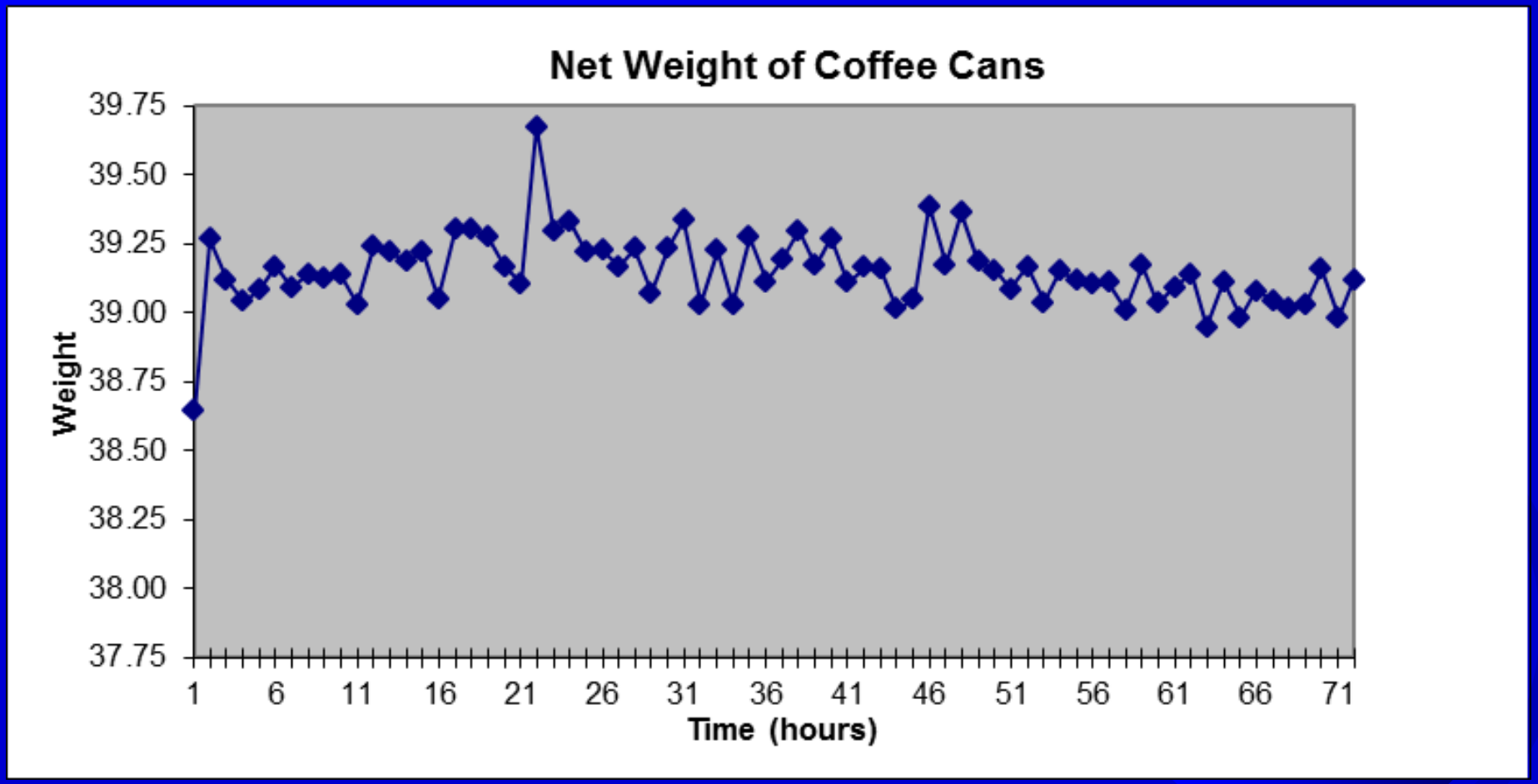
Common Features of Time Series

Time series have time-dependent features:

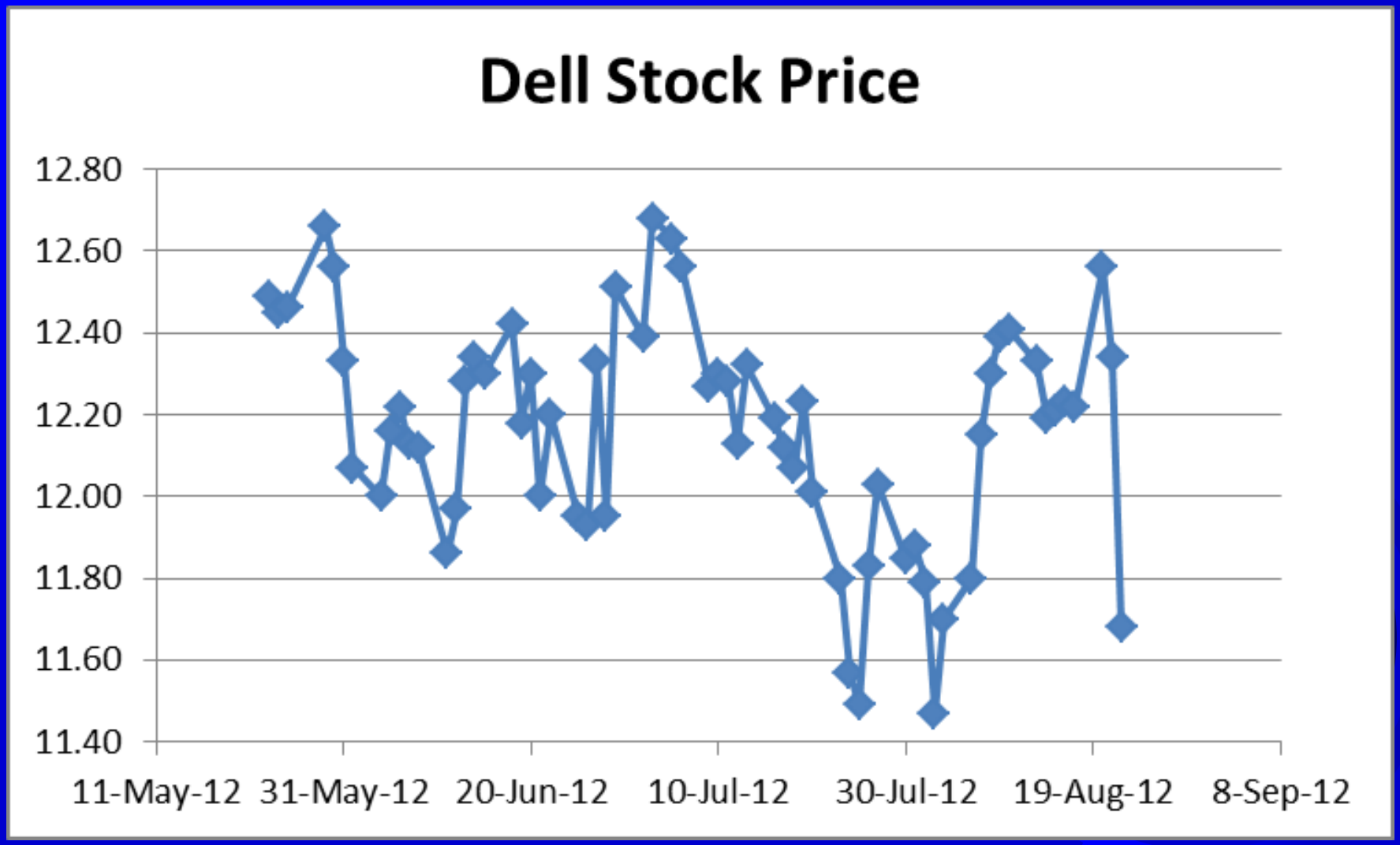
- Trends
- Seasonal patterns
- Autocorrelation (dependence on own past)
- Cross-correlation (dependence on other time series)

A Real Time Series

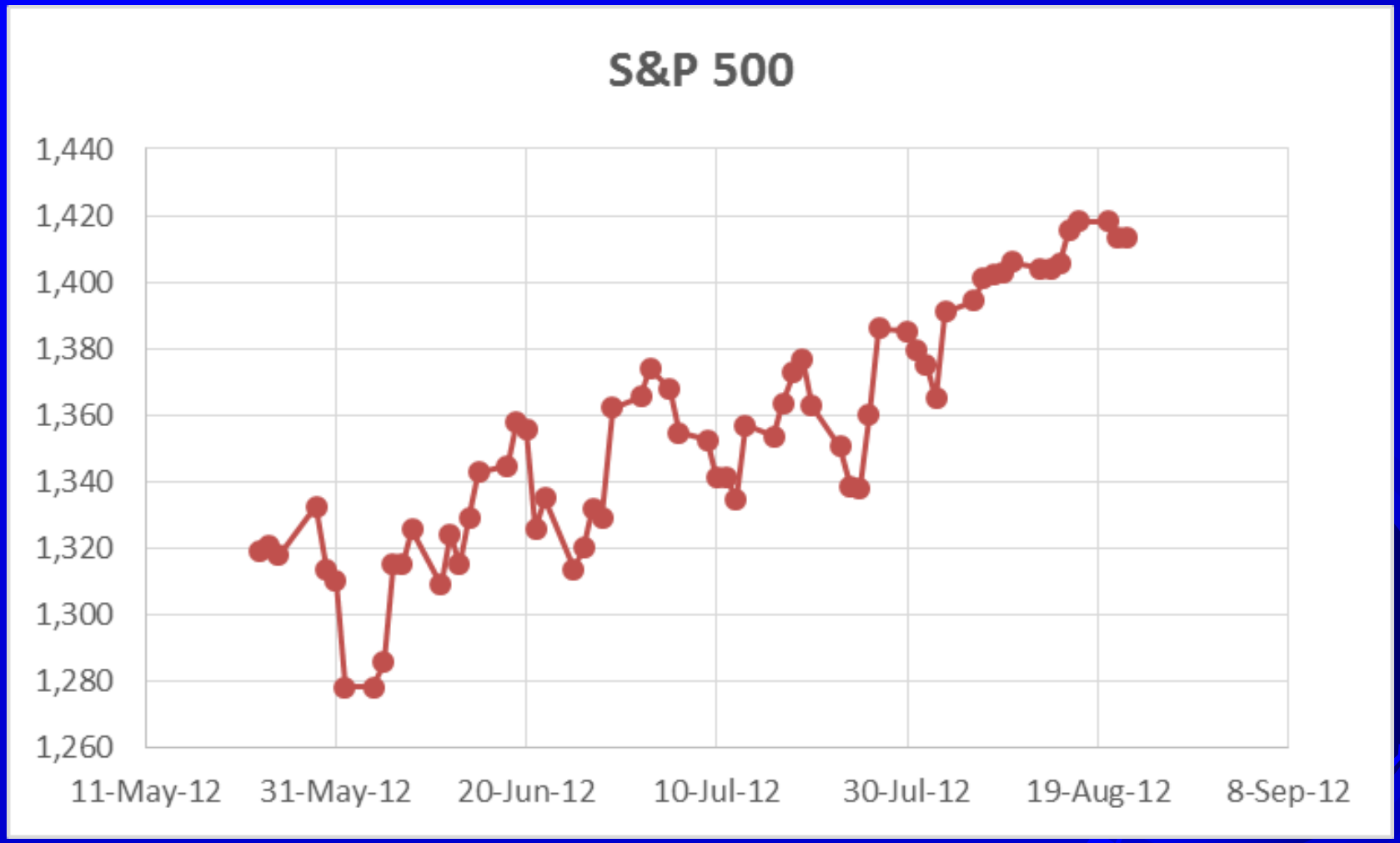
A Real Time Series



Another Real Time Series



A Related Time Series



Why study time series?

- To forecast

  - to predict future values of the time series

- To explain

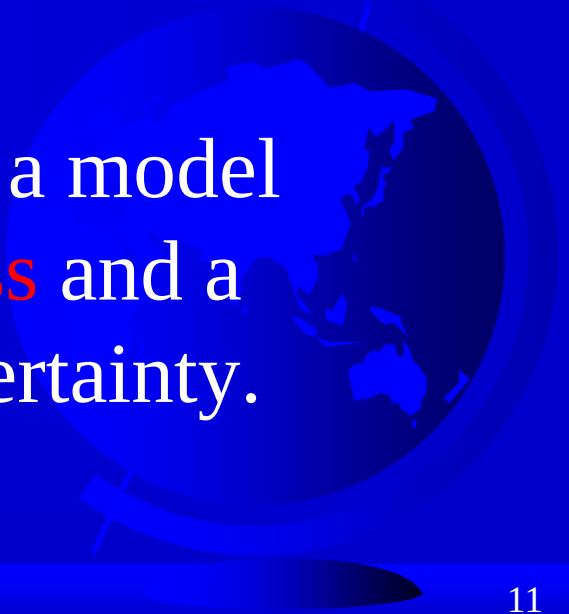
  - to understand the factors responsible for the values of the time series

“*Prediction is very difficult, especially if it's about the future.*”

- Niels Bohr

Essence of Forecasting

- Forecasting relies on the premise that the past foreshadows the future
- Every good forecast starts with a simplified model of the past that can be extended into the future
- Every good forecast is based on a model and consists of a numerical guess and a numerical assessment of its uncertainty.



Good Models Capture the Important Features of Time Series

- Trends
- Seasonal patterns
- Autocorrelation (dependence on own past)
- Cross-correlation (dependence on other time series)

Examples of Time Series Models

- Random Sample
- Random Walk
- Autoregression
- Moving Average
- ARIMA (Autoregressive Integrated Moving Average)
- Panel data
- Vector Autoregression (VAR)

Key Modeling Steps

1. Propose
2. Validate
3. Use

Descriptive Statistics

Mean, Standard deviation, Correlation

- ***Mean***

  How big is a typical data value?

- ***Standard deviation***

  How spread out are the data?

- ***Correlation***

  How strong is the relationship between two paired sets of data?

Descriptive Statistics

Mean and Standard Deviation

- ***Mean***

  Excel: =average(*range*)

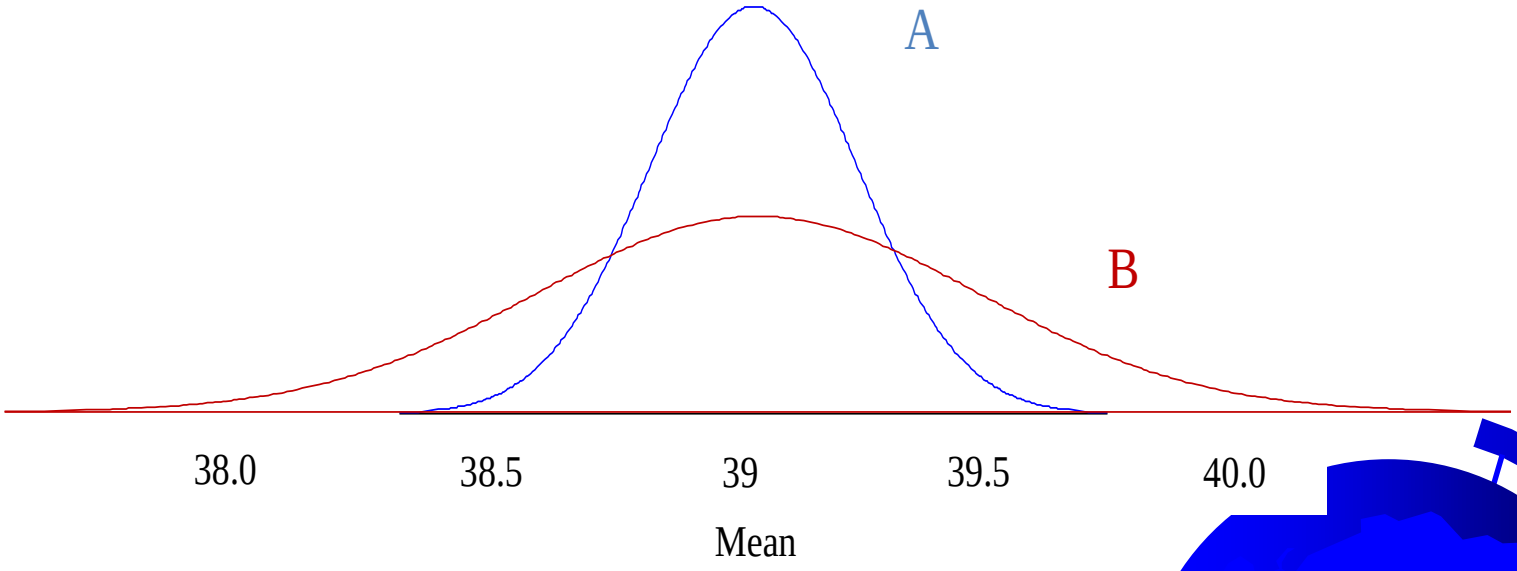
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- ***Standard deviation***

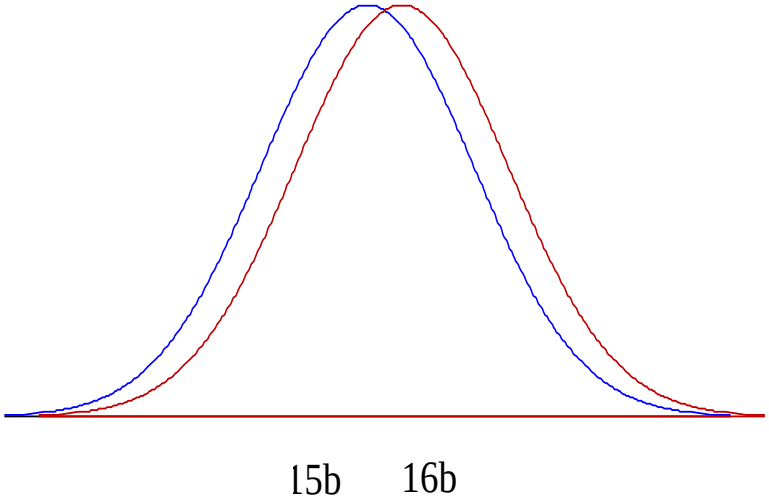
  Excel: =stdev(*range*)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

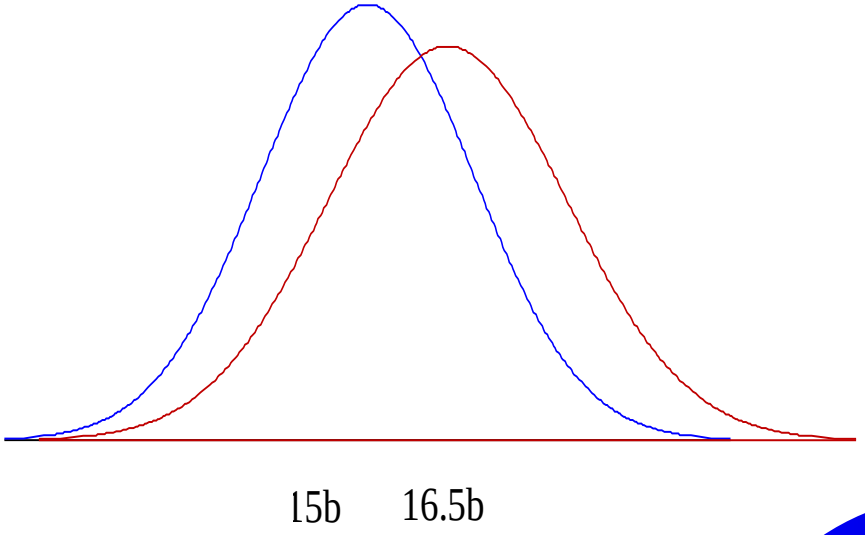
Same Mean, Different StDev



Different Mean, Same StDev



Different Mean, Different StDev



Descriptive Statistics

Correlation Coefficient

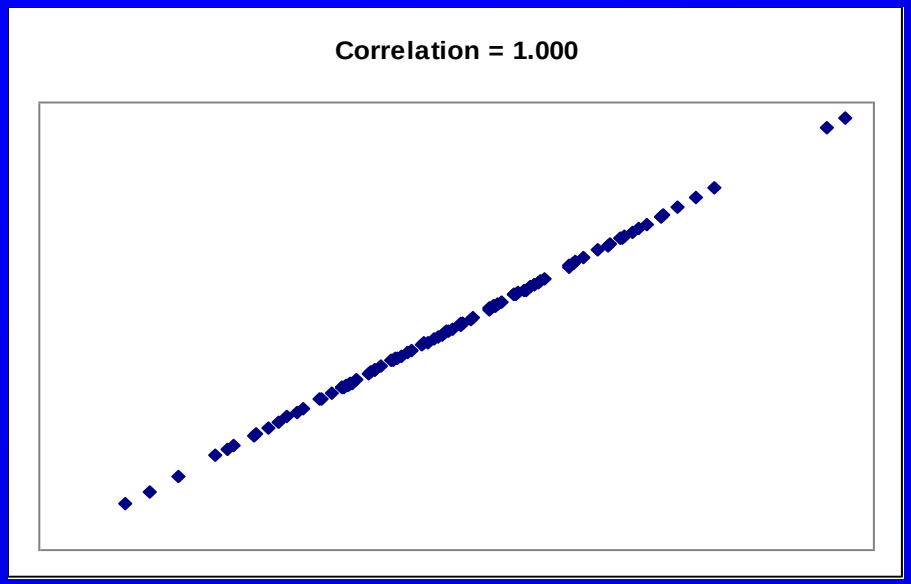
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

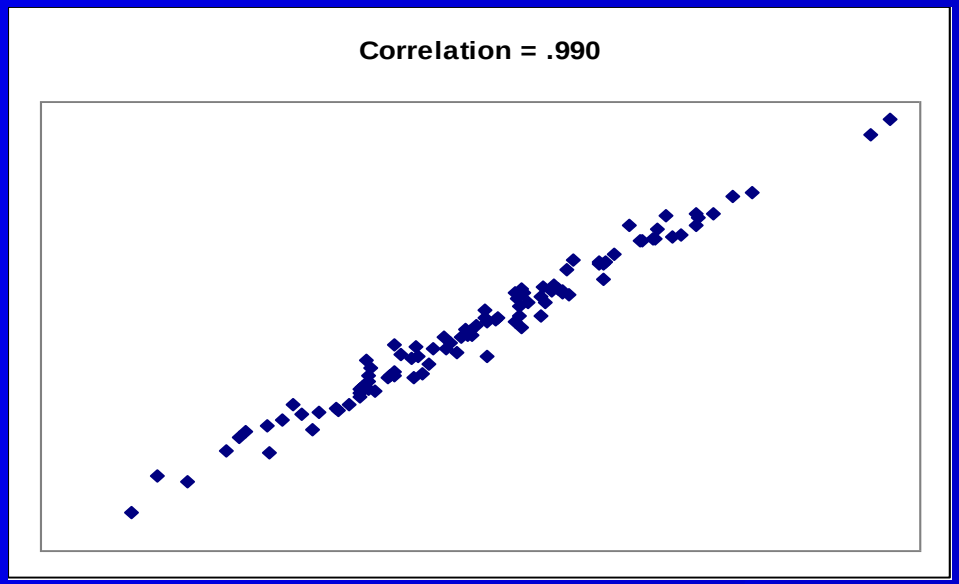
- Excel: =correl(*Xrange*,*Yrange*)

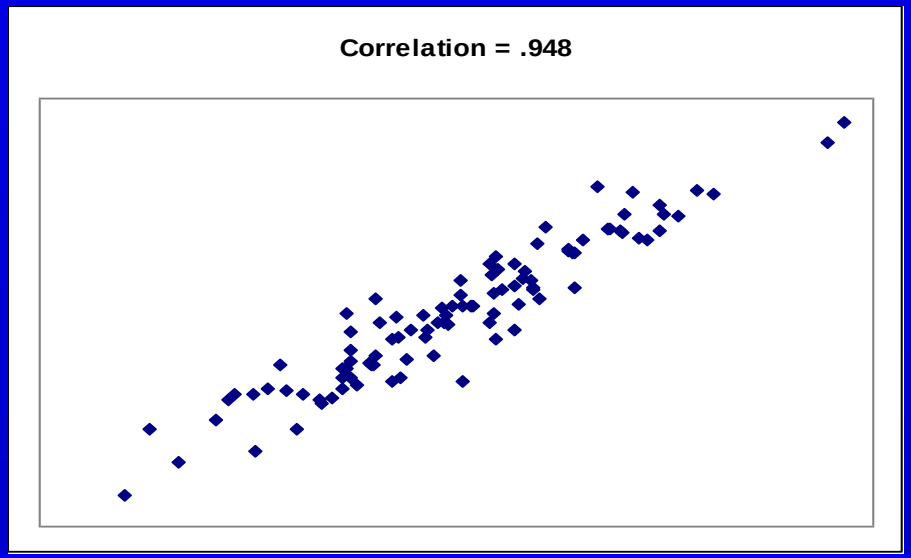
Properties:

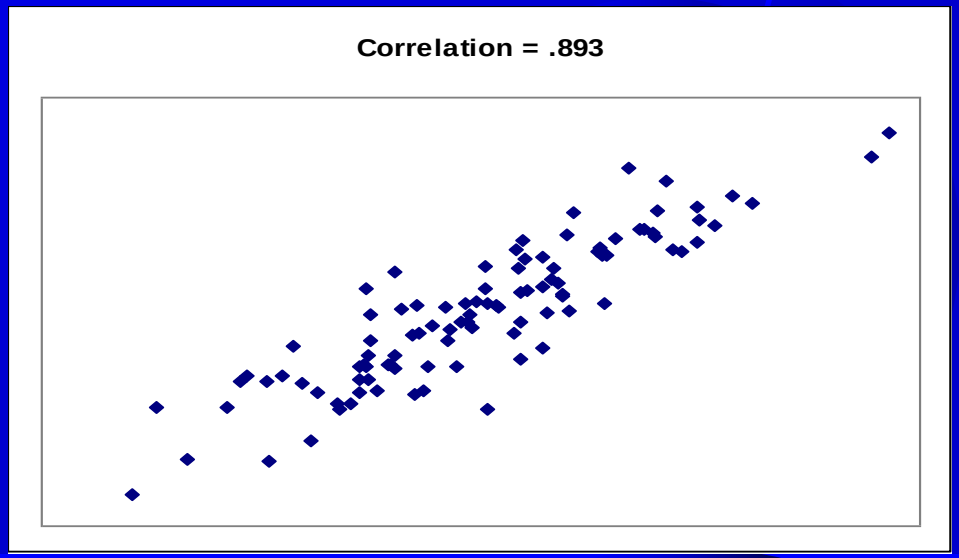
1. -1 $\leq$ r $\leq$ +1
2. r = +1 means perfect positive relationship
3. r = -1 means perfect negative relationship
4. r = 0 means no (linear) relationship

Correlation Examples

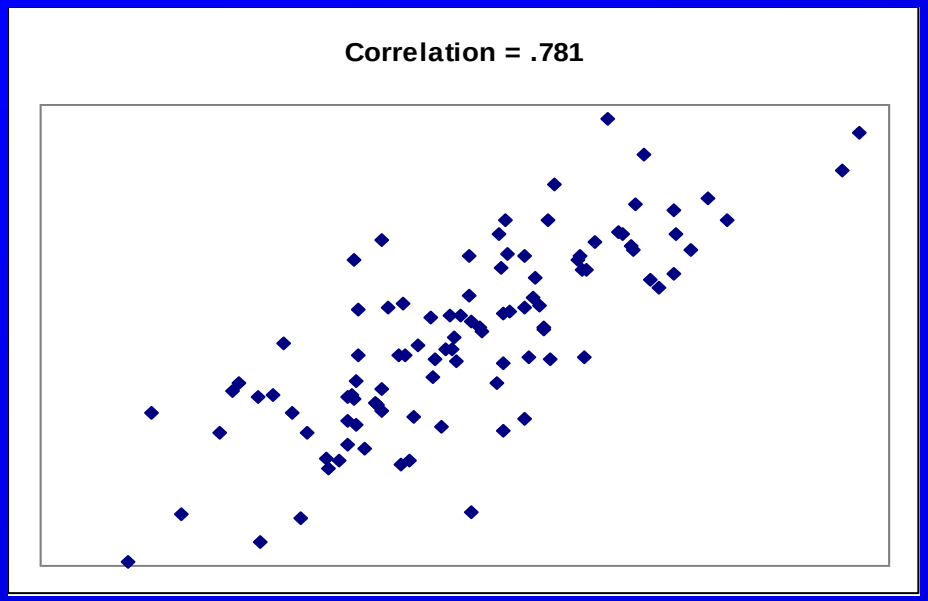


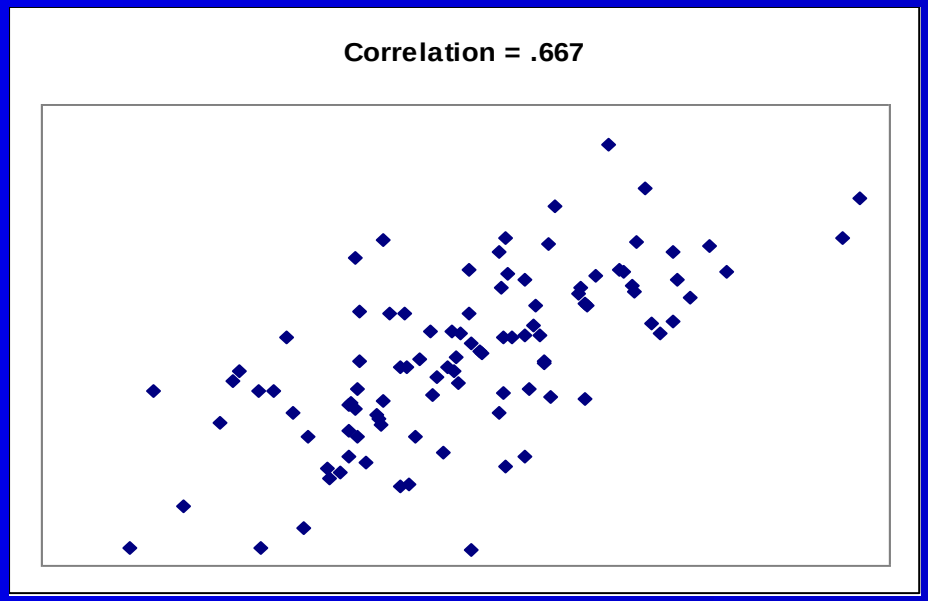


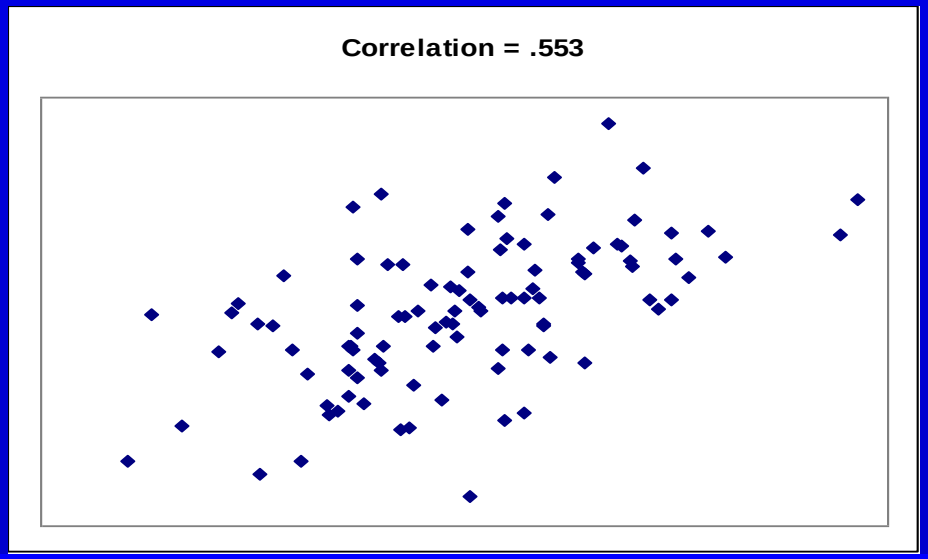


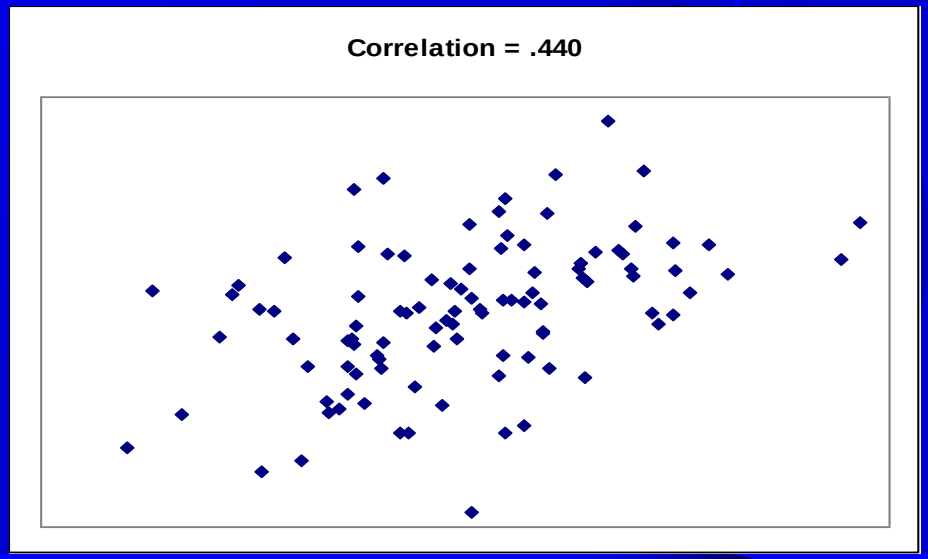


More Correlation Examples

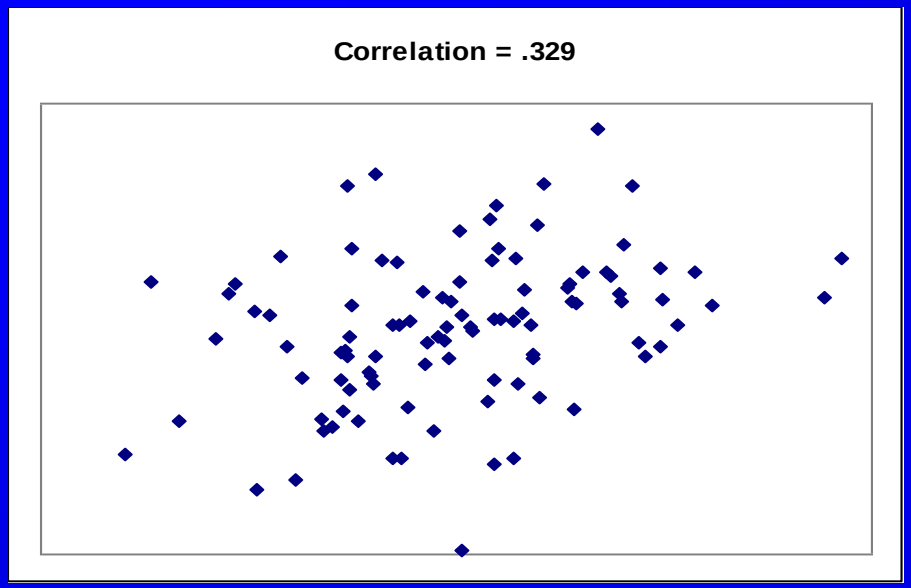


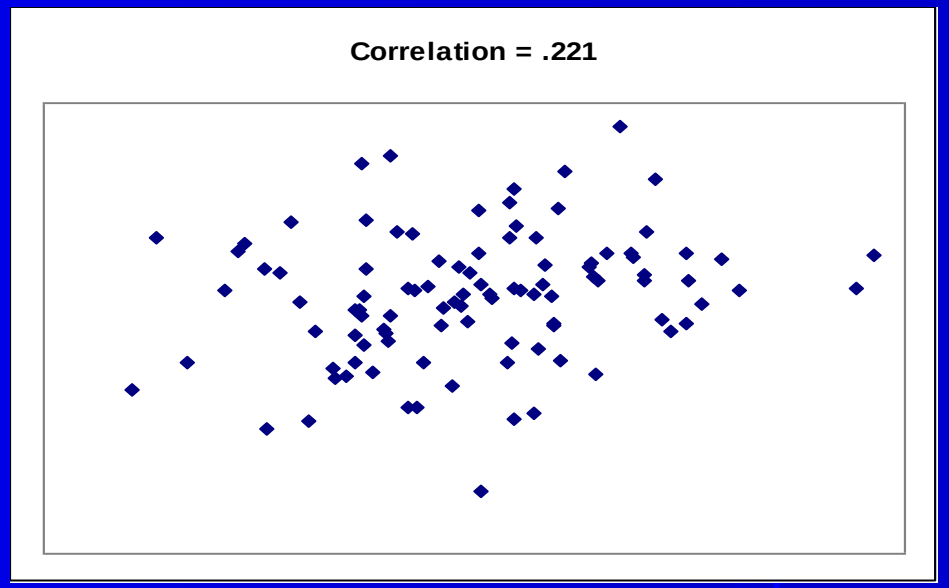


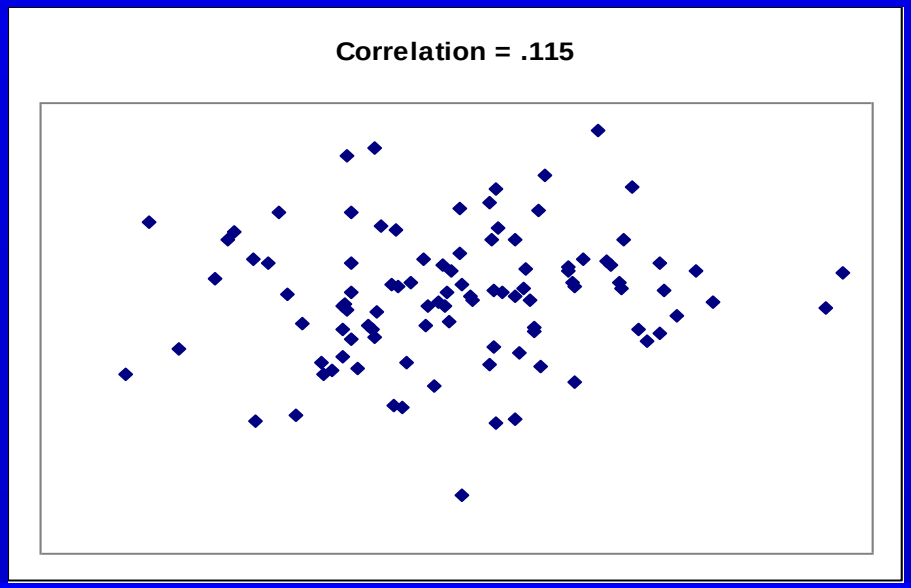


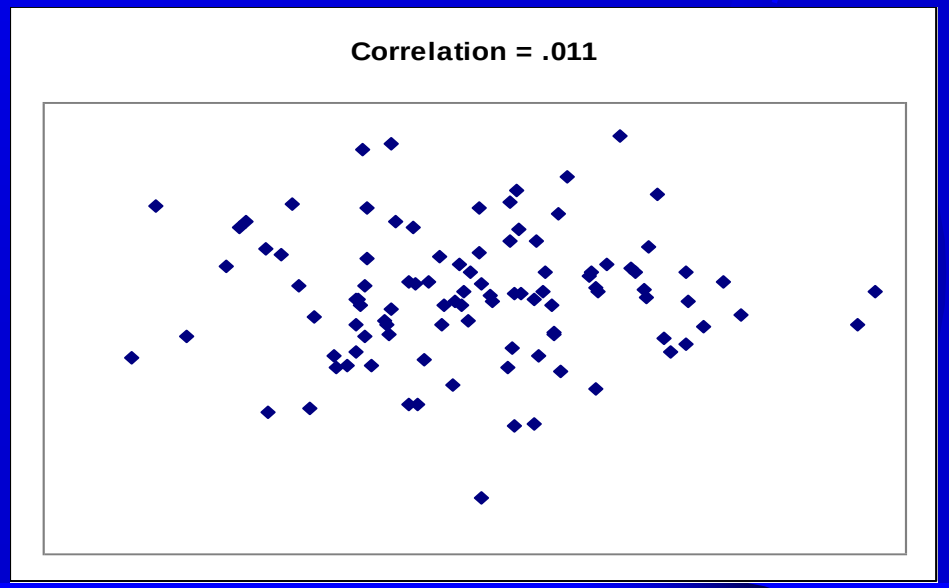


Still More Correlation Examples

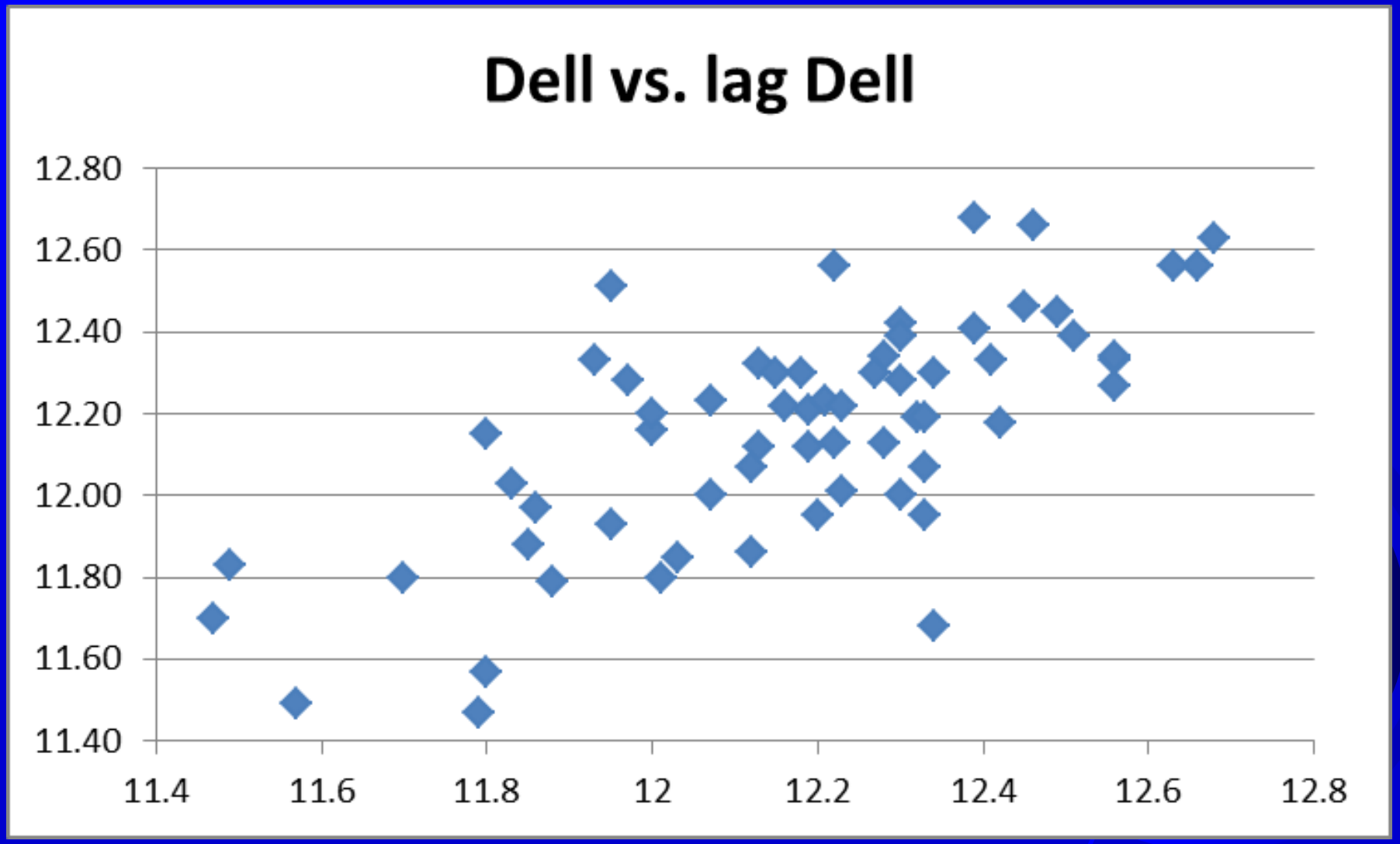






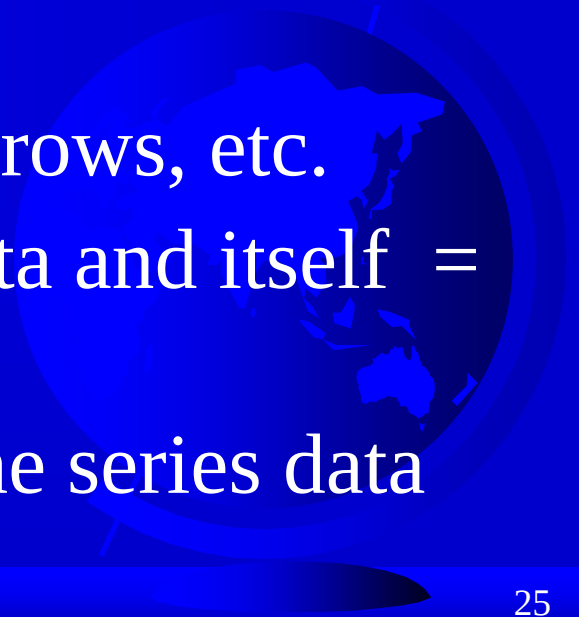


Correl(*Dell, lag Dell*) = 0.70

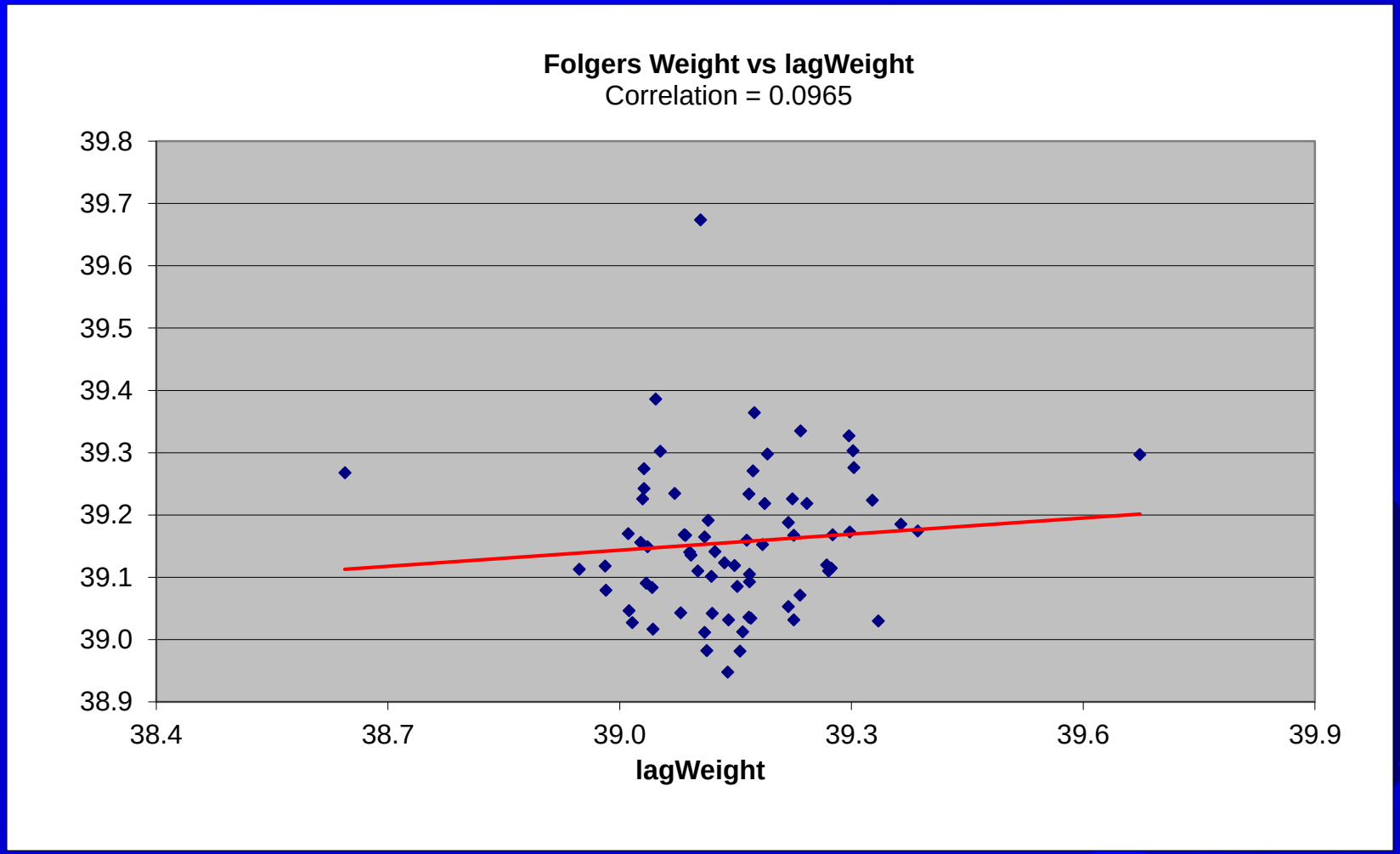


Autocorrelation

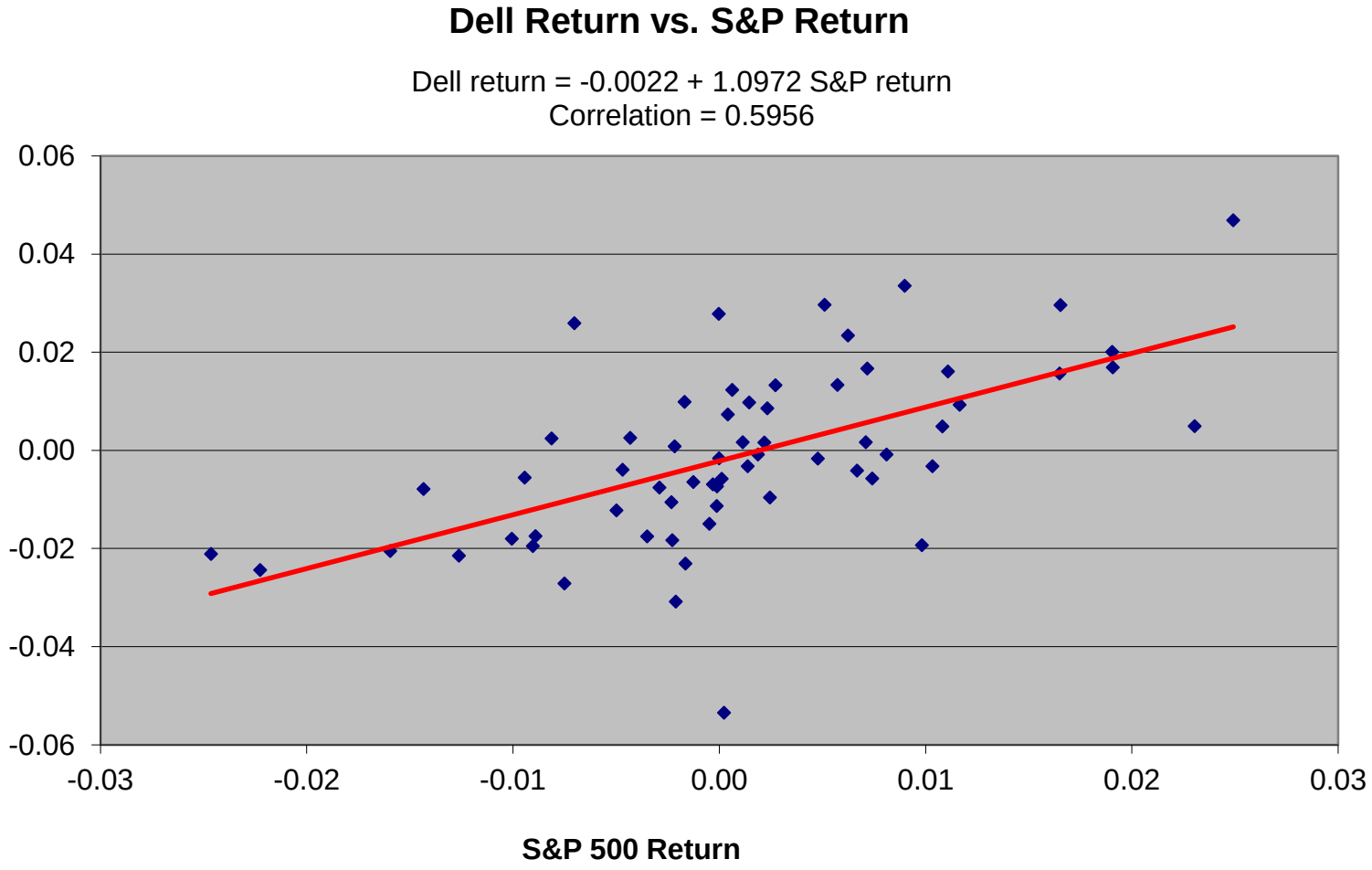
- Tells how time series data depend on their predecessors in time (e.g., today's price on yesterday's, this year' sales on last year's, etc.)
- Calculate as correlation between a column of data and itself, ***offset by one row***
- This is *lag 1* autocorrelation
- Lag 2 autocorrelation offsets by 2 rows, etc.
- Correlation between column of data and itself = +1.00
- Autocorr makes sense only for time series data



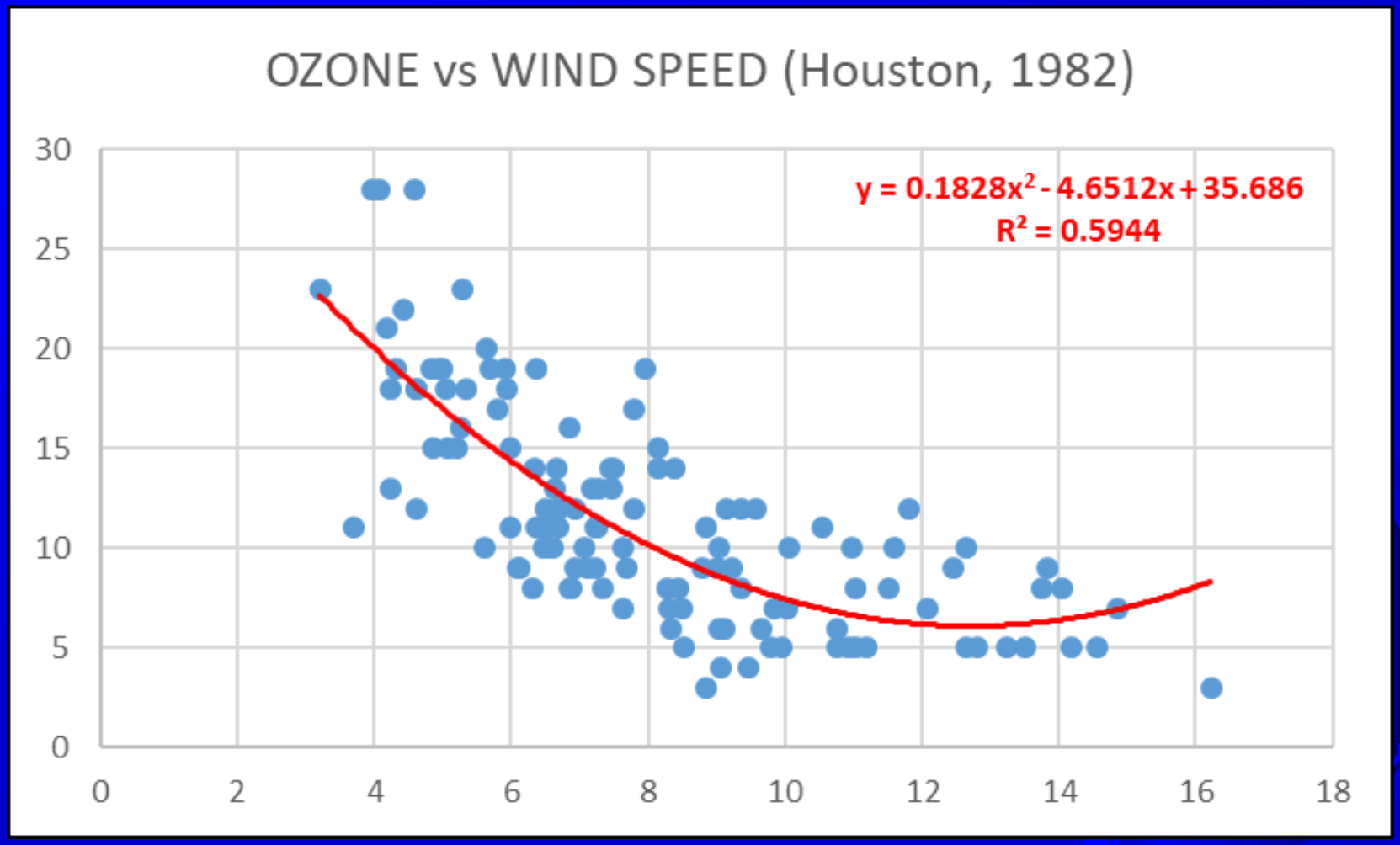
Correl(*Weight, lag Weight*) = 0.0965



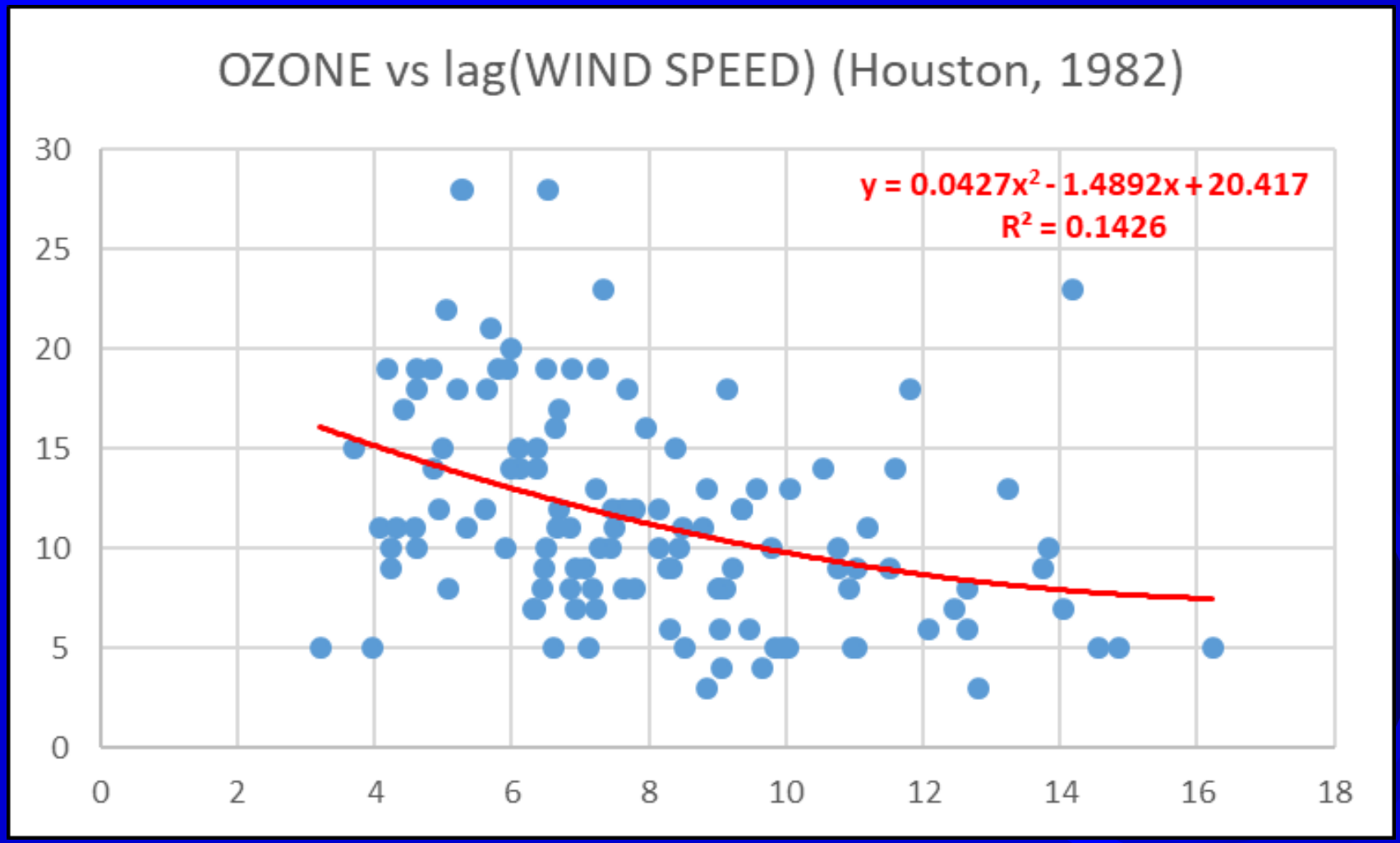
Correl(*Dell Return, S&P Return*) = 0.5956



Modeling

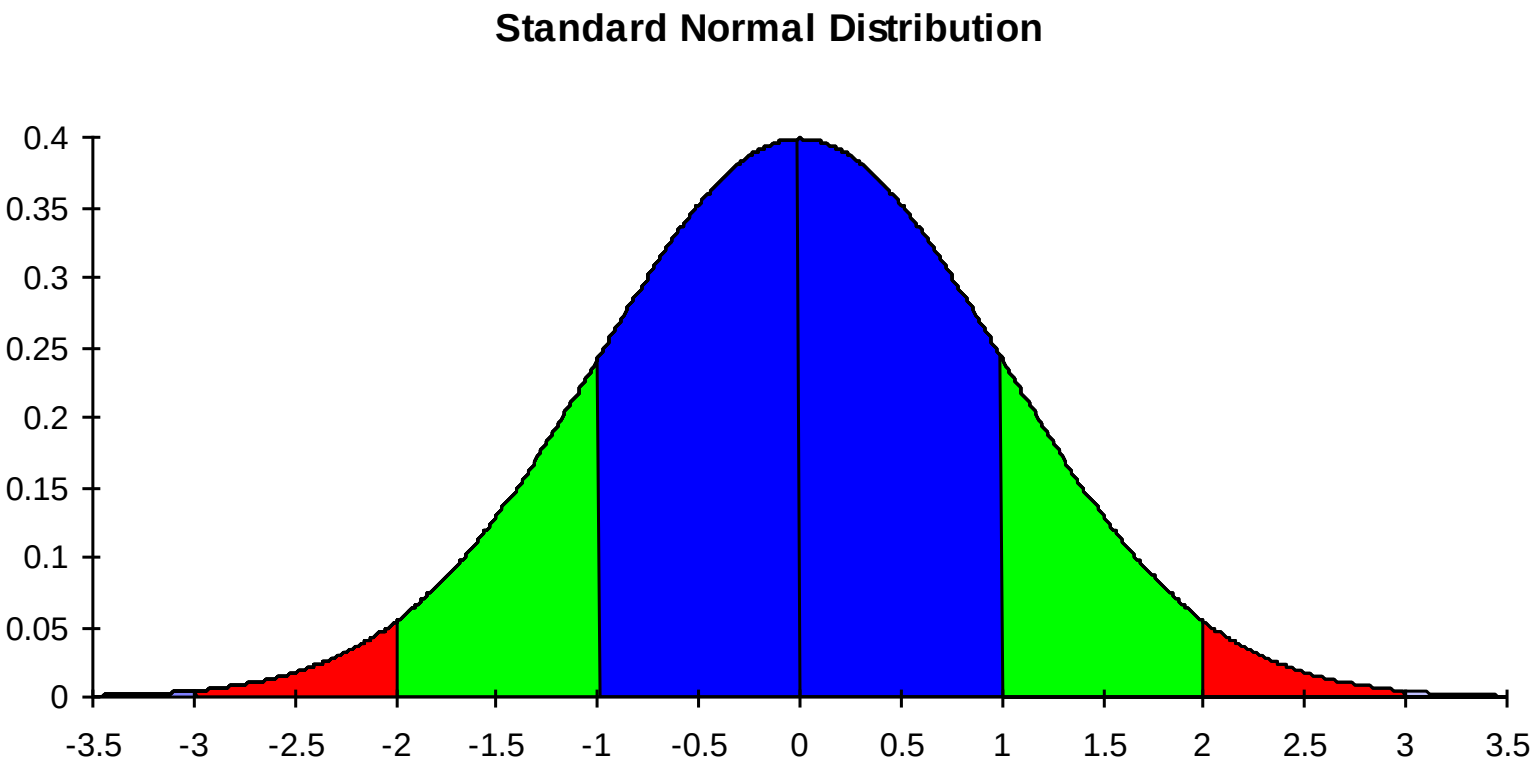

OZONE vs WIND SPEED (Houston, 1982)
y = 0.1828x² - 4.6512x + 35.686
R² = 0.5944
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18

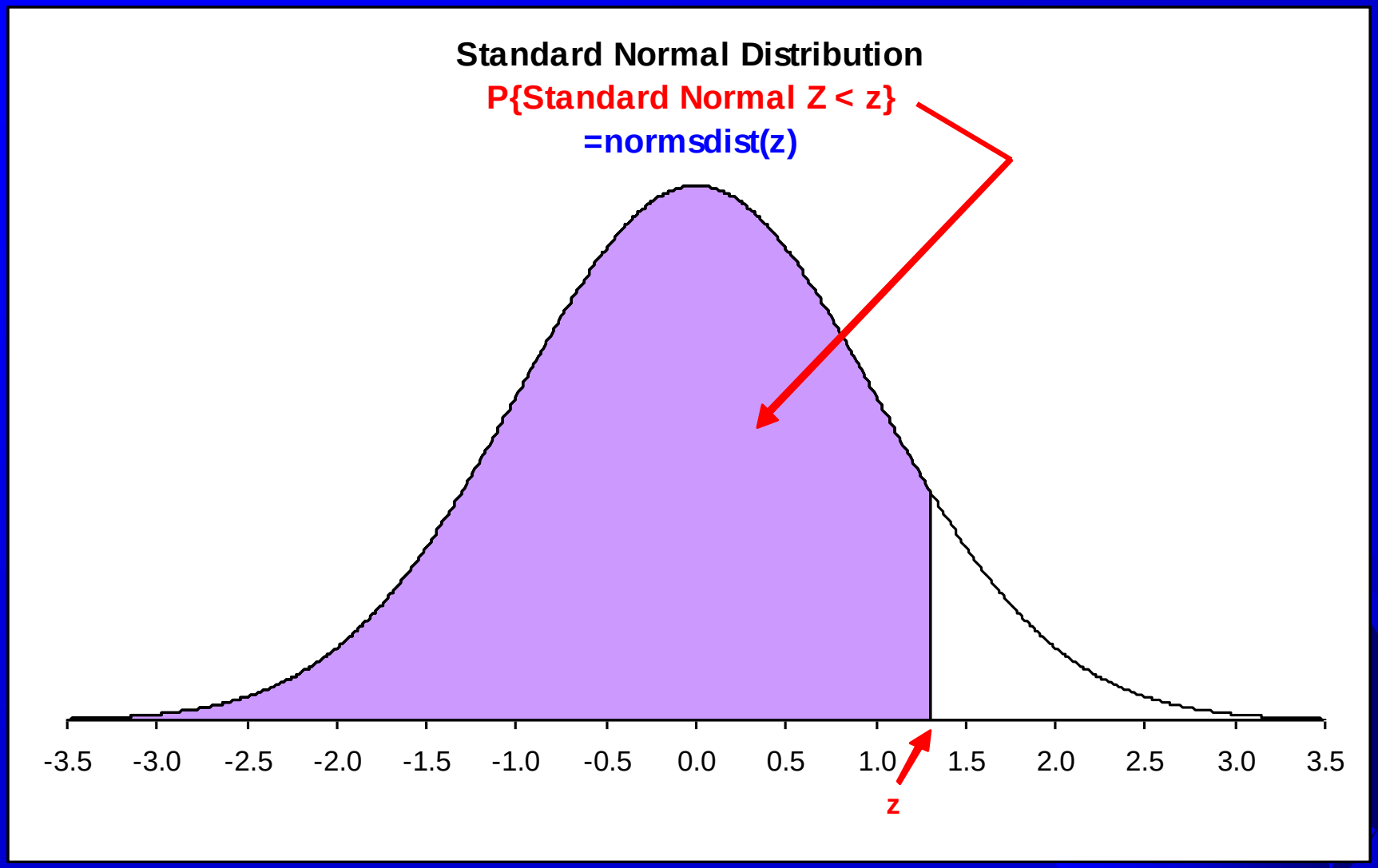
Modeling


OZONE vs lag(WIND SPEED) (Houston, 1982)
y = 0.0427x² - 1.4892x + 20.417
R² = 0.1426
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18

Standard Normal Distribution

Mean = 0, StDev = 1




Standard Normal Distribution
P{Standard Normal Z < z}
=normsdist(z)
-3.5
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
z