Longitudinal Analysis

CSE545 - Fall2017

Supplemental Presentation

Introduction Time Series Analysis

Goal: Understanding temporal patterns of data (or real world events)

Common tasks:

- **Trend Analysis:** Extrapolate patterns over time (typically descriptive).
- **Forecasting:** Predicting a future event (predictive).
  (contrasts with "cross-sectional" prediction -- predicting a different group)

Introduction to Causal Inference (Revisited)

X causes Y as opposed to X is associated with Y

Changing X will change the distribution of Y.

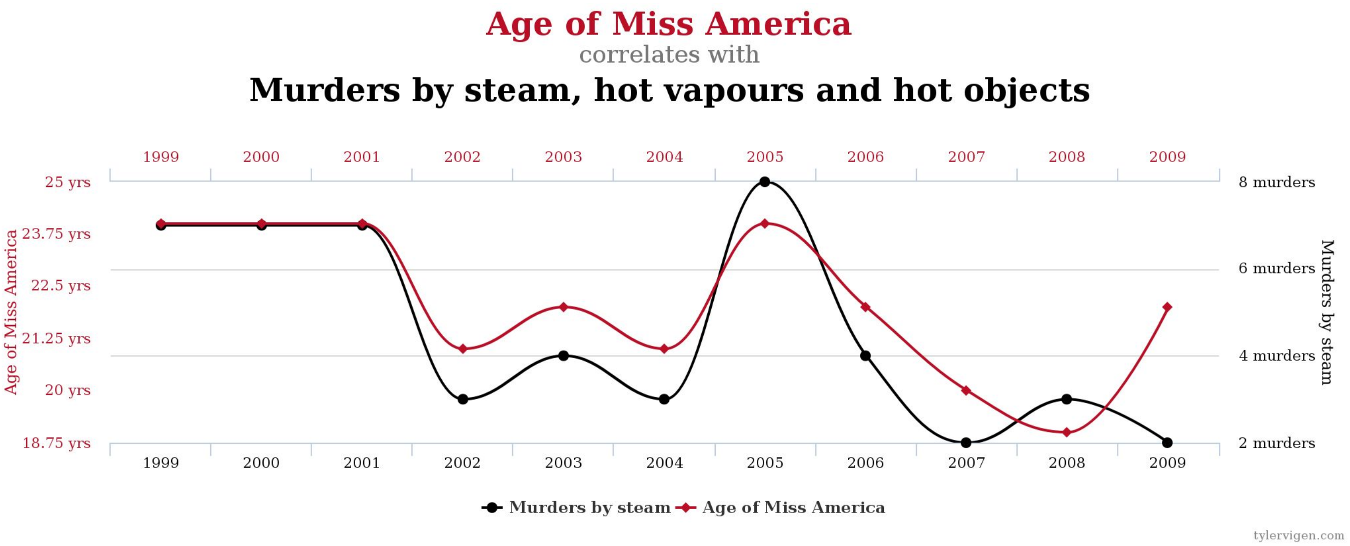
X causes Y $\not\Leftrightarrow$ Y causes X

Spurious Correlations

Extremely common in time-series analysis.

Spurious Correlations

Extremely common in time-series analysis.



Introduction to Causal Inference (Revisited)

X causes Y as opposed to X is associated with Y

Changing X will change the distribution of Y.

X causes Y $\nleftrightarrow$ Y causes X

$$P(Y = 1|X = 1) - P(Y = 1|X = 0)$$

Counterfactual Model: Exposed or Not Exposed: X = 1 or 0

$$Y = \begin{cases} C_0 & \text{if } X = 0 \\ C_1 & \text{if } X = 1 \end{cases}$$

Causal Odds Ratio:

$$\frac{\left(\frac{P(C_1=1)}{P(C_1=0)}\right)}{\left(\frac{P(C_0=1)}{P(C_0=0)}\right)}$$

Autocorrelation

“(a.k.a. Serial correlation).”

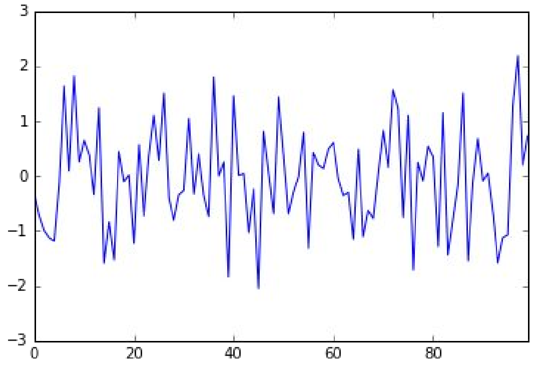
Quantifying the strength of a temporal pattern in serial data.

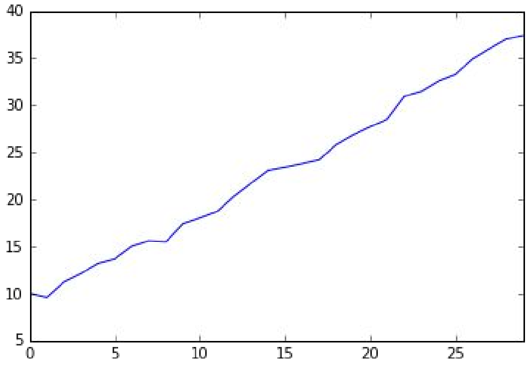
Requirements:

- Assume regular measurement (hourly, daily, monthly...etc..)

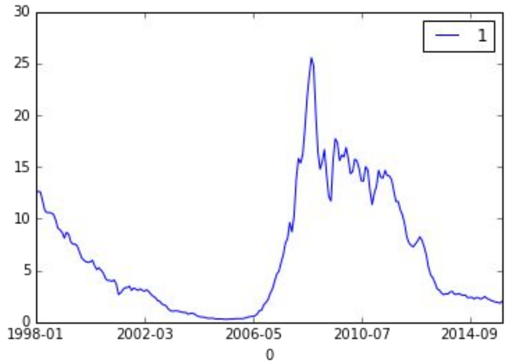
Autocorrelation

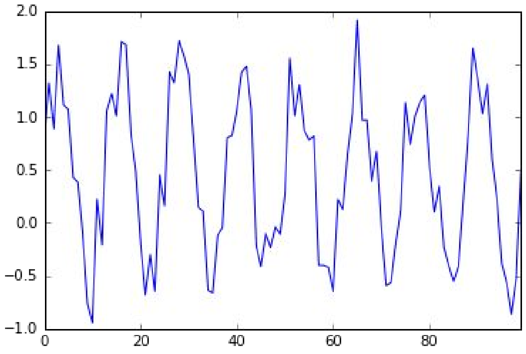
Quantifying the strength of a **temporal pattern** in serial data.





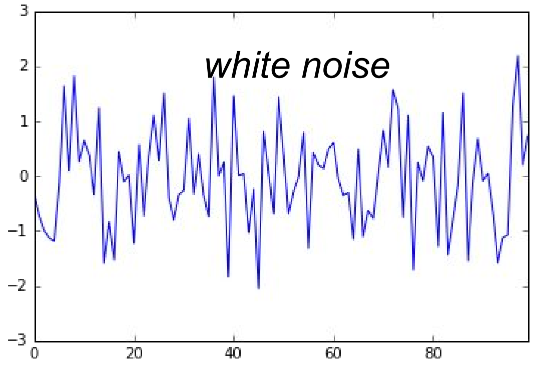
Which have temporal patterns?

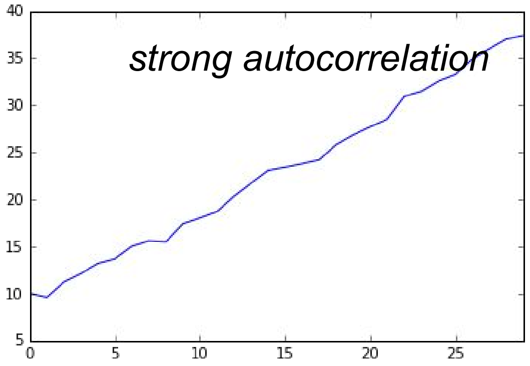




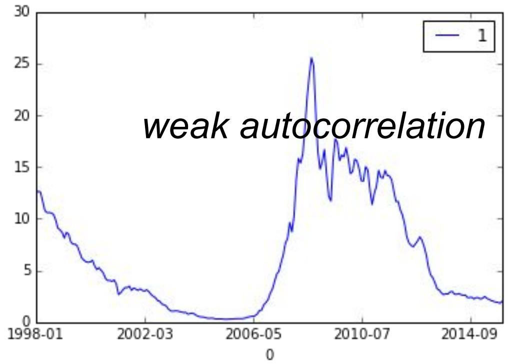
Autocorrelation

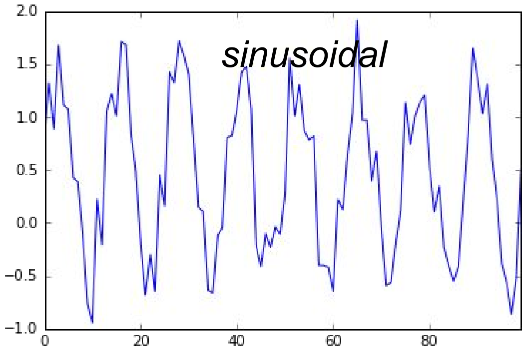
Quantifying the strength of a **temporal pattern** in serial data.





Which have temporal patterns?

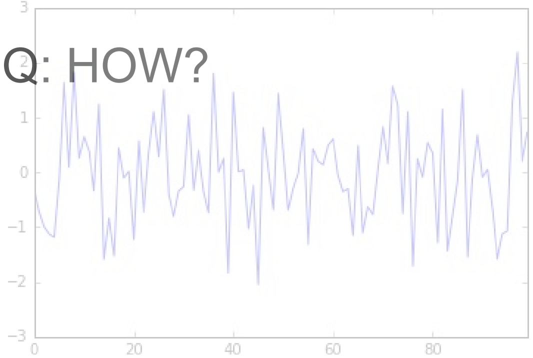


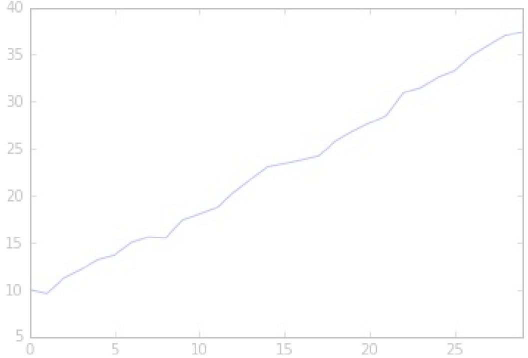


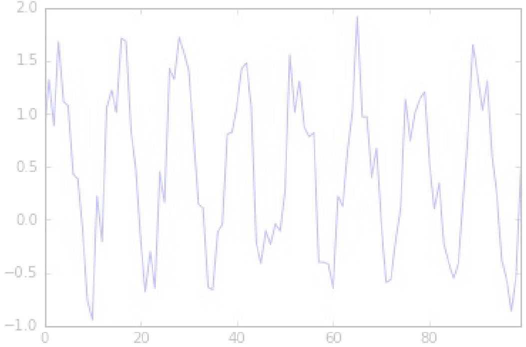
Autocorrelation

Quantifying the strength of a **temporal pattern** in serial data.

Q: HOW?







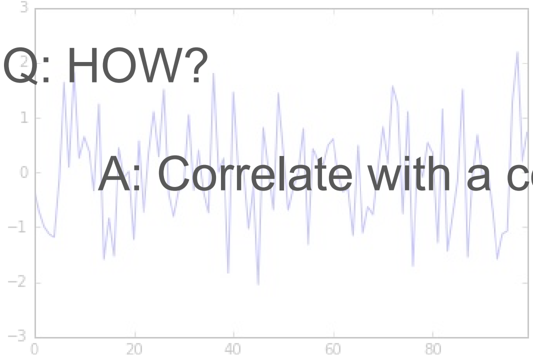
Autocorrelation

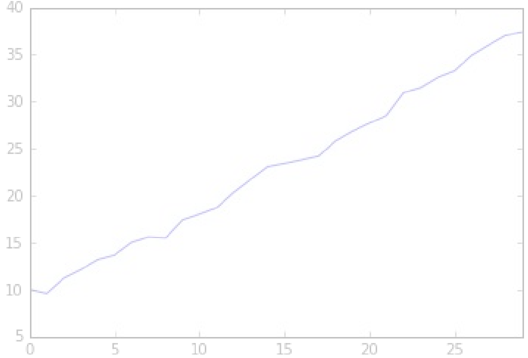
Quantifying the strength of a **temporal pattern** in serial data.

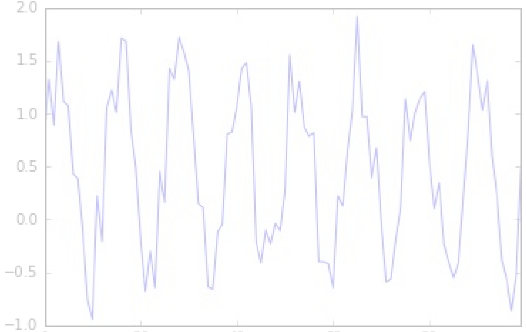
Q: HOW?

A: Correlate with a copy of self, shifted slightly.

….







Autocorrelation

Quantifying the strength of a **temporal pattern** in serial data.

Q: HOW?

A: Correlate with a copy of self, shifted slightly.

```
Y = [3, 4, 4, 5, 6, 7, 7, 8]

correlate(Y[0:7], Y[1:8])  #lag=1

correlate(Y[0:-2], Y[2:8])  #lag=2

....
```


Autocorrelation

Quantifying the strength of a **temporal pattern** in serial data.

Q: HOW?

A: Correlate with a copy of self, shifted slightly.

```
Y = [3, 4, 4, 5, 6, 7, 7, 8]

correlate(Y[0:7], Y[1:8])  #lag=1

correlate(Y[0:-2], Y[2:8])  #lag=2

....
```


Autoregressive Model

AR Models: $$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, ..., Y_{t-n}, \epsilon_t)$$

Linear AR model: $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_n Y_{t-p} + \epsilon_t$$

Autoregressive Model

AR Models: $$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, ..., Y_{t-n}, \epsilon_t)$$

Linear AR model: $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_n Y_{t-p} + \epsilon_t$$

Notation:
$$\begin{aligned}
\mathrm{AR}(1)&: \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} \\
\mathrm{AR}(2)&: \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} \\
\mathrm{AR}(3)&: \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}
\end{aligned}$$

Autoregressive Model

AR Models: $Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, ..., Y_{t-n}, \epsilon_t)$

Linear AR model: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_n Y_{t-p} + \epsilon_t$

Notation:

$$\text{AR(1)}:\ \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1}$$
$$\text{AR(2)}:\ \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2}$$
$$\text{AR(3)}:\ \hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3}$$

$$\text{AR(0)}:\ \hat{Y}_t = \beta_0$$

Moving Average

Based on error;   (a “smoothing” technique).

Q: Best estimator of random data (i.e. white noise)?

Moving Average

Based on error;   (a “smoothing” technique).

Q: Best estimator of random data (i.e. white noise)?

A: The mean

$$\hat{Y}_t^{MA} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \ldots + Y_{t-p}}{p+1}$$

Moving Average

Based on error;   (a "smoothing" technique).

Q: Best estimator of random data (i.e. white noise)?

A: The mean

$$\hat{Y}_t^{MA} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \ldots + Y_{t-p}}{p+1}$$

Simple Moving Average

Moving Average Model

In a regression model (AR**MA** or ARI**MA**), we consider error terms

$$Y_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \ldots)$$

Moving Average Model

In a regression model (AR**MA** or ARI**MA**), we consider error terms

$$Y_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \dots)$$

$$\hat{Y}_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_p\epsilon_{t-p}$$

Moving Average Model

In a regression model (AR**MA** or ARI**MA**), we consider error terms

$$Y_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \ldots)$$

$$\hat{Y}_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_p \epsilon_{t-p}$$

attributed to "shocks" -- independent, from a normal distribution

Notation:

$$\text{MA}(1)\text{: } \hat{Y}_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$$

$$\text{MA}(2)\text{: } \hat{Y}_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

ARMA Models

AutoRegressive (AR) Moving Average (MA) Model

ARMA(p, q):

$$\hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

ARMA(1, 1):

$$\hat{Y}_t = \beta_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$$

example: Y is sales; error may be effect from coupon or advertising

(credit: Ben Lambert)

ARIMA

I = Integrated

Makes a time series stationary:

- Removes trends (“detrending”)
- Makes “mean reverting” = tendency to always revert back to the mean over the long run.
- Removes changes in variance

Time-series Applications

- ARMA
  - Economic indicators
  - System performance
  - Trend analysis
    (often situations where there is a general trend and random “shocks”)
- Univariate Models in General
  - Anomaly Detection
  - Forecasting
  - Season Trends
  - Signal Processing

- Integration as predictors within multivariate models

statsmodels.tsa.arima_model