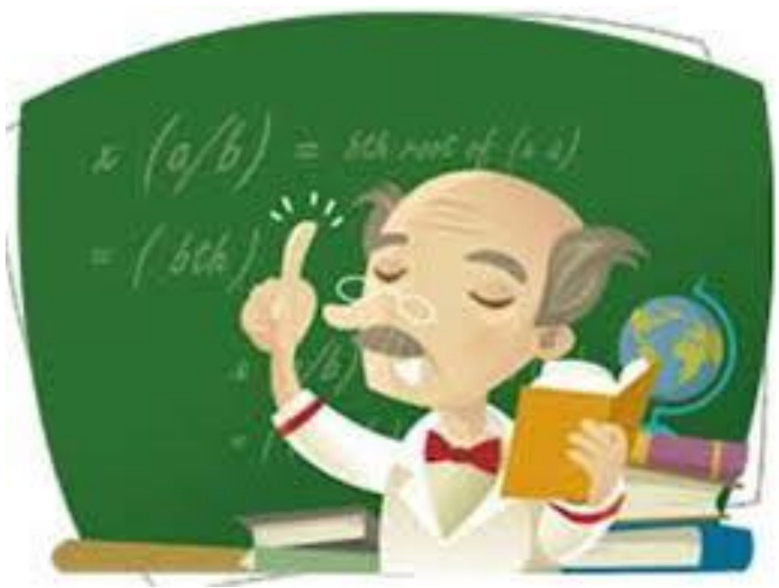
PSTAT 174/274: Intro to Time Series
Week 1: Autocorrelations Stationarity



Instructor: Raya Feldman

Note: Raya as in

PSTAT 174/274: Intro to Time Series
Week 1: Autocorrelations Stationarity

Outline of Lecture 1

Outline of Lecture 2

Time Series PSTAT 174/274: Syllabus Points

Textbooks:

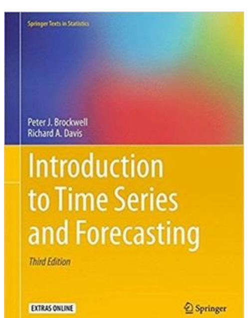
- No required textbook.
- Class lectures are self-contained
- **Lecture notes/ slides will be posted on the class website on Friday of the previous week**
- **Complement lectures by reading**

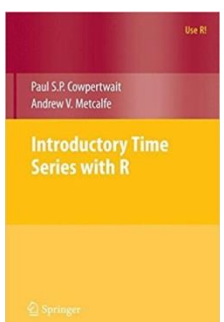
➢ *Introduction to Time Series and Forecasting*, P. Brockwell and R. Davis, Springer

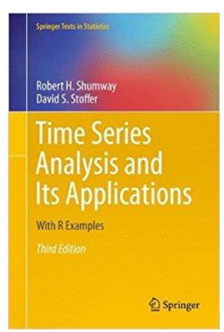
➢ *Time Series Analysis with R Examples*, R. H. Shumway and D. S. Stoffer, Springer. For free e-textbook see UCSB Library.
Check Professor Stoffer's website http://www.stat.pitt.edu/stoffer/tsa4/

CAS recommendation for Time Series portion of Exam MAS-I:

➢ *Introductory Time Series Analysis with R* by P. S. P. Cowpertwait and A.V. Metcalfe, Springer. The pdf of the book is available for free from UCSB Library.







Question:
Are ALL Times Series Books Yellow?

More Syllabus Points: Grading

Prerequisites: PSTAT 10, PSTAT 120A-B with a letter grade C or better
Highly recommended: PSTAT 126 (Regression)

Grading:

5% Lab participation (two lowest grades dropped)
15% Homework (the lowest grade dropped)
30% Quizzes (the average of four quiz grades; two attempts for each quiz)
50% Final project (individual)

Labs:

- Attend your lab section in person to work with TA and other students
- lab assignments graded P/NP, submitted via Gradescope;
- to receive a pass, least 75% or the assignment must be completed;
- posted Tuesday morning, due Tuesday noon next week.

Quizzes and Homework:

- offered through Gaucho Space /Gradescope, see separate file with information

Academic Honesty:

- No identical pieces of homework, lab assignment, quizzes or final project.
- No identical R code or text. No submitting work available on the Web as your own.
- Read more at https://studentconduct.sa.ucsb.edu/academic-integrity

More Syllabus Points: PSTAT 174 vs. 274

Office hours and Communication (tentative): Zoom,email, Nectir

Office hours and instructor/TAs emails posted on GauchoSpace

Links to Zoom and Nectir are provided on Gaucho Space

PSTAT 274: additional assignments, stricter grading of final project.

Reasons to move to PSTAT 274:

- **More solid knowledge of the material**
- **Honors students get honors units eliminating need for honors contracts**
- **Students considering combined BS/MS Act Sci program**
- **Students planning to apply to graduate school**

UG students are allowed to enroll into a graduate class if

(i) **A student has a minimum cumulative GPA of 3.0, and**

(ii) **A student has at least 12 units of upper-division credit with a grade of B or better in PSTAT or related courses, e.g., Math.**

Procedure for enrolling into PSTAT 274 is posted in a separate file.

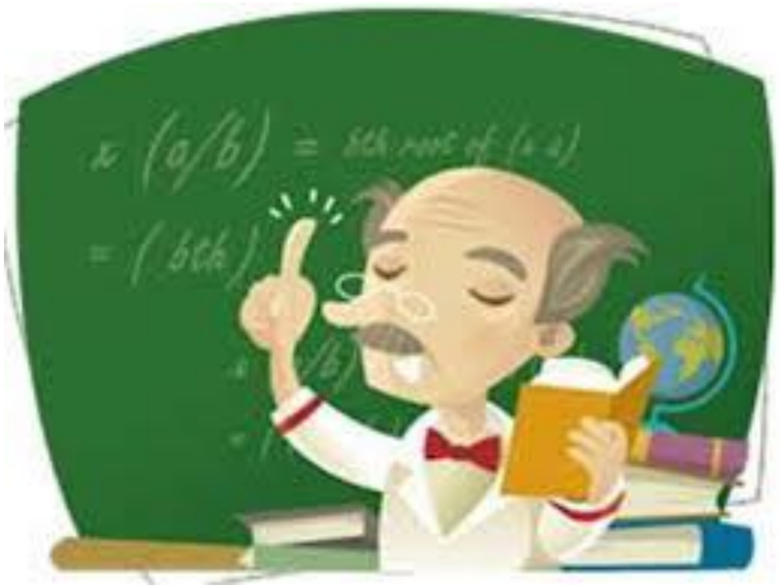
But what is Time Series?

PSTAT 174/274: Intro to Time Series

Lecture 1, Part 2 : Time Series Data

Part 2 of Lecture 1, Outline:

- Characteristics of Time Series Data
- Examples of Time Series Data
- Applications of Time Series
- Goals of Time Series Analysis



But what is Time Series?
Ask Mr. Google?

What is Time Series?

Mr. Google:

There is (time series) data everywhere

Time Series are an old idea, the city of Barcelona stores data about its citizens extensively since 13th Century

From https://mapr.com/blog/time-series-data-new-big-data/

What Time Series Are

or

How is Time Series Data Different

from all Other Data ?

What Time Series Are

or

How is Time Series Data Different from all Other Data ?

Hints: In most of your courses

- **The data points were pretty much independent**
- **The order in which observations (data points) came in not important (cross-sectional data)**

What Are Time Series or
How is Time Series Data Different ?

Hints: In most of your courses

- **The data points were pretty much independent**
- **The order in which observations (data points) came in where not important (cross-sectional data)**

- **Time Series Data: a series of values recorded in time**
- **Order: very important! Observations dependent!**

Question: Can you give an example of a time series?

- **Time Series Data: a series of values recorded in time**
- **Order: very important! Observations dependent!**

Examples:

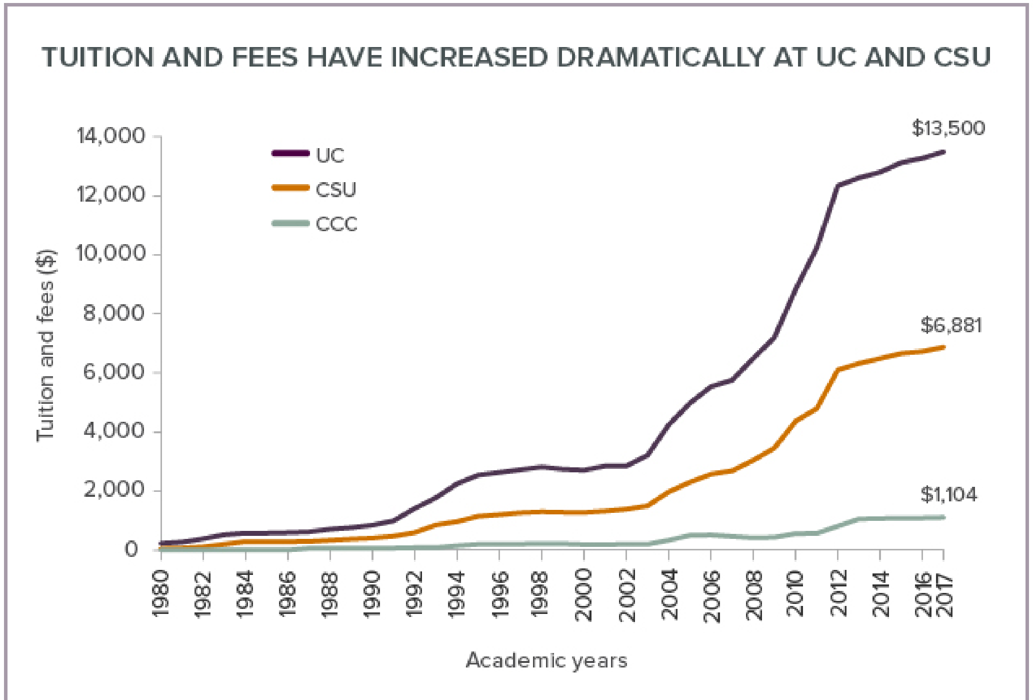
class enrollments in PSTAT 174 per quarter;

population of US per year;

daily sales in Costco, etc.

Main Question asked:

- **forecast the future with somewhat reliable error estimate**



SOURCE: University of California, Office of the President; California State University Chancellor's Office; and California Community Colleges Chancellor's Office.
NOTE: Adjusted for inflation, in 2016 dollars. Tuition at the California Community Colleges (CCC) is for a full-time load of 12 units each semester and does not include campus fees.

https://www.ppic.org/higher-education/affordability/

Is It Time Series ?

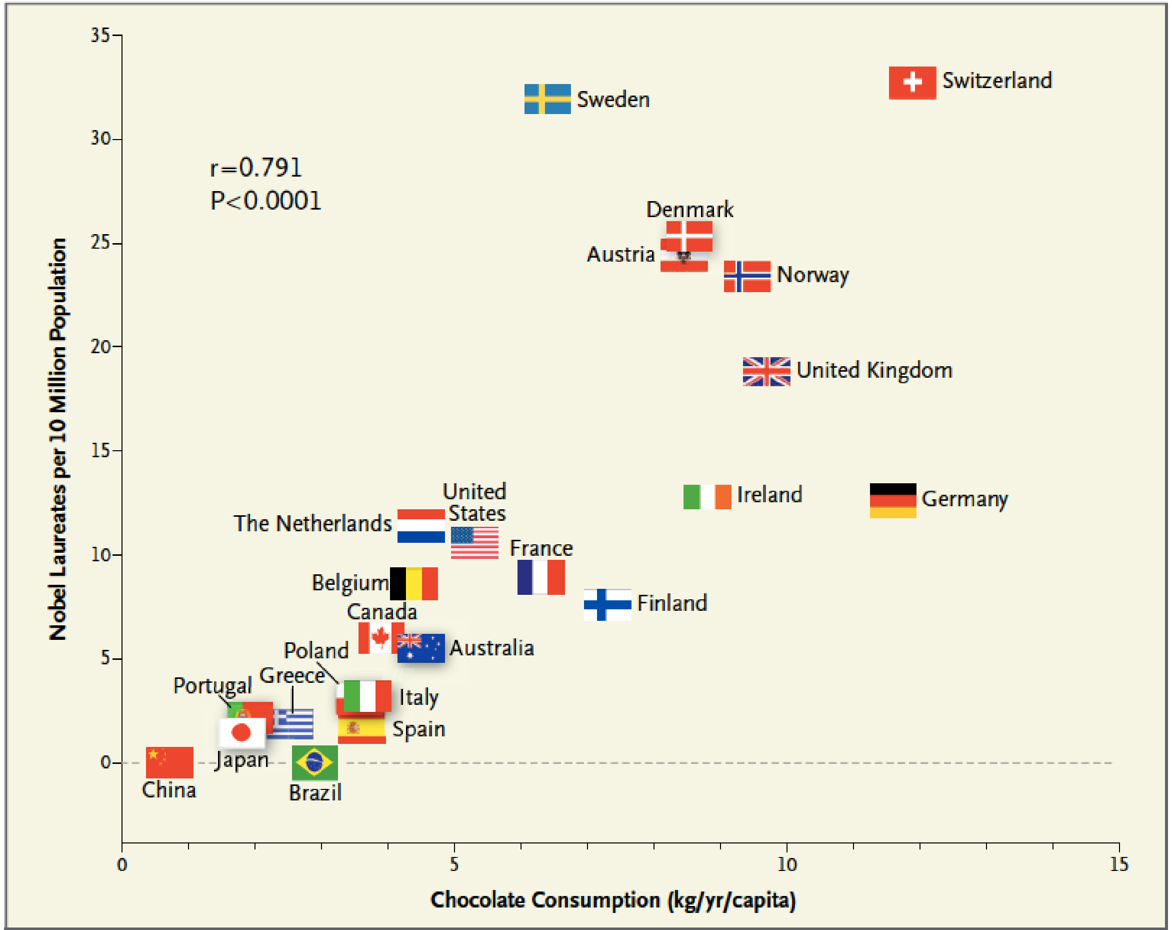


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

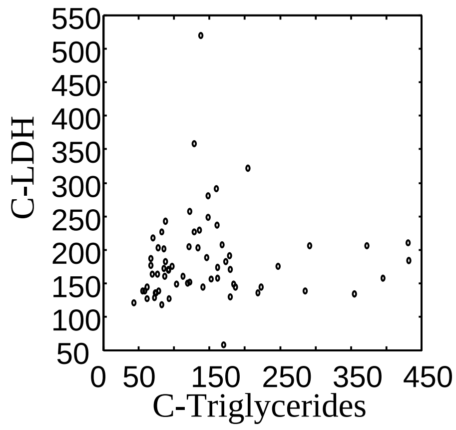
Independent bivariate data (x, y) = (chocolate consumption, Nobel prize winners)
Order of observations does not matter (cross-sectional data)

from http://www.businessinsider.com/chocolate-consumption-vs-nobel-prizes-2014-4?r=UK&IR=T

Example from a Machine Learning class: Large number of features

- **Example: 53 Blood and urine measurements (wet chemistry) from 65 people (33 alcoholics, 32 non-alcoholics).**

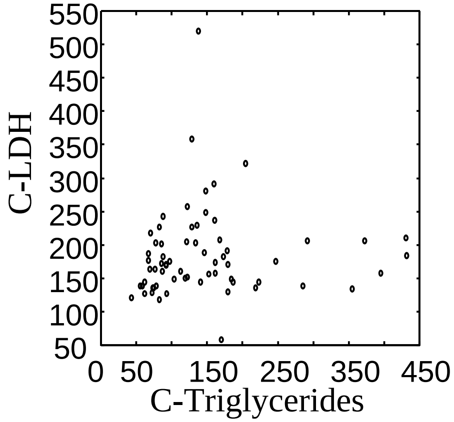
	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000



Example from a Machine Learning class: Large number of features

- **Example: 53 Blood and urine measurements (wet chemistry) from 65 people (33 alcoholics, 32 non-alcoholics).**

	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000



Is It Time Series ?

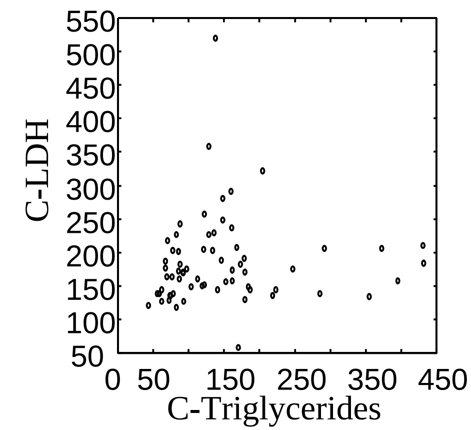
Ask: Is Order Important?

Ask: Are Observations independent?

Example from a Machine Learning class: Large number of features

- **Example: 53 Blood and urine measurements (wet chemistry) from 65 people (33 alcoholics, 32 non-alcoholics).**

	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000



Is It Time Series ?

?

Is Order Important?

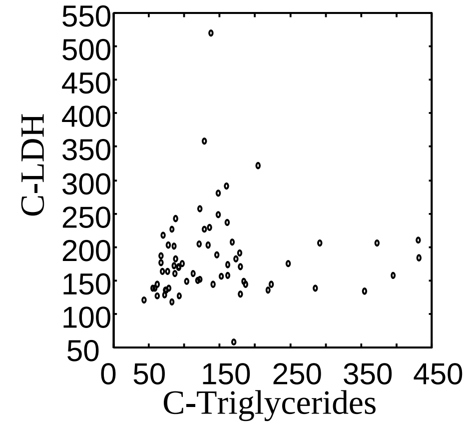
Are Observations independent?

- **Order is unimportant**
- **Independent data points (A1, A2, etc.), each 53-dimensional**
- **Complicated dependence between multivariate variables (H-WBC, H-RBC, etc.);**

Example from a Machine Learning class: Large number of features

- **Example: 53 Blood and urine measurements (wet chemistry) from 65 people (33 alcoholics, 32 non-alcoholics).**

	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000



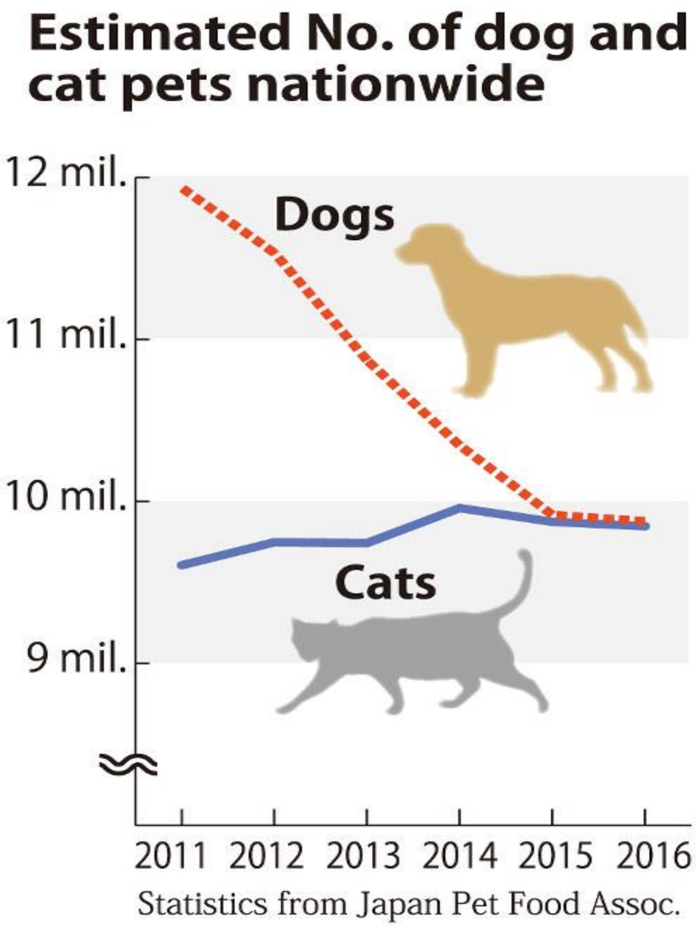
- **Order is unimportant**
- **Independent data-points (A1, A2, etc.)**
- **Complicated dependence between multivariate variables (H-WBC, H-RBC, etc.);**

Questions asked:

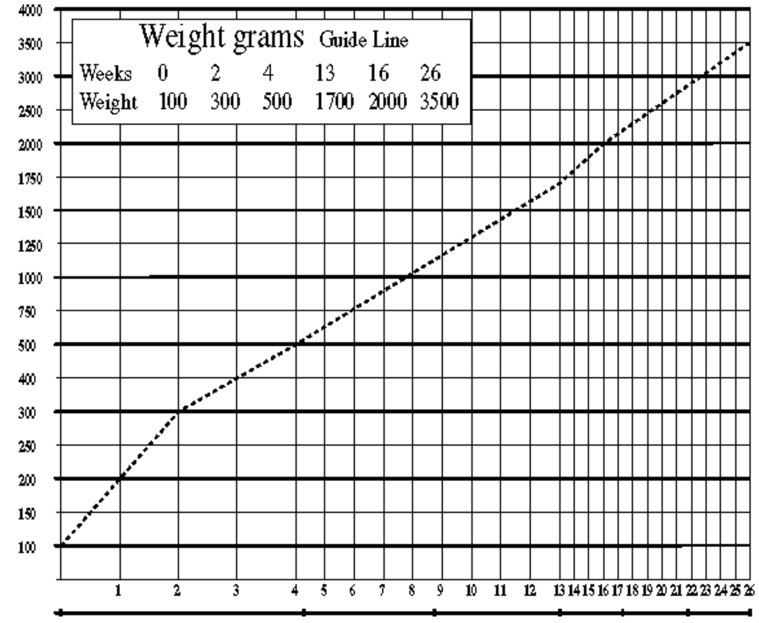
- **Do we need a 53 dim space to view data?**
- **How to find the 'best' low dim space that conveys maximum useful information?**
- **One answer: Find "Principal Components"**

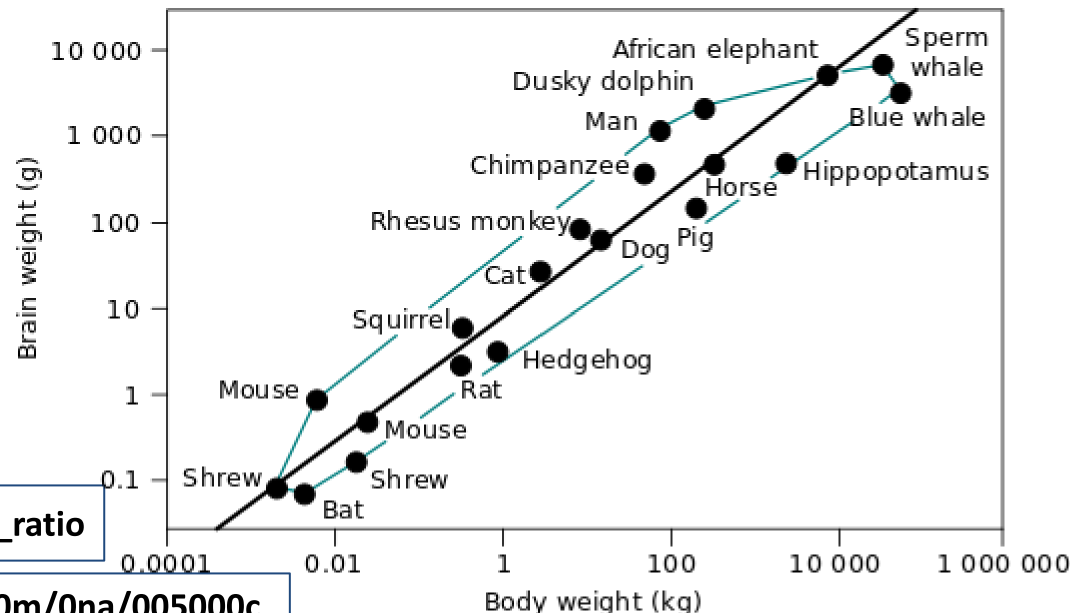
Not Time Series

3 examples: Is it Time Series Data?



Weekly Kitten Weight:





https://en.wikipedia.org/wiki/Brain-to-body_mass_ratio

https://mainichi.jp/english/articles/20170118/p2a/00m/0na/005000c

http://shorthair.candra.info/domestic-shorthair-cat-weight-chart.html

- **Time Series Data: a series of values recorded in time**

- **Most common: equally spaced time-points:**

  $X_t, X_{t+h}, X_{t+2h}, X_{t+3h}, \ldots,$ **or** $X(t), X(t+h), X(t+2h), \ldots$

h= time between observations or *sampling interval;*

1/h = *sampling frequency* or *sampling rate*

- **Order: very important! Observations dependent!**
- **Notation: position in the series, not actual time:** $X_1, X_2, X_3, \ldots, X_n$

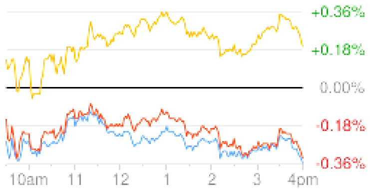
- **Goal of time series: forecast**

- **Examples: class enrollments in PSTAT 174 per quarter;**
  population of US per year;
  daily sales in Costco, etc.

Time series forecasting is performed in nearly every organization that works with quantifiable data:

- **Retails stores use it to forecast sales.**
- **Energy companies use it to forecast reserves, production, demand, and prices.**
- **Educational institutions use it to forecast enrollment.**
- **International financial organizations such as the World Bank and International Monetary Fund use it to forecast inflation and economic activity.**
- **Transportation companies use time series forecasting to forecast future travel.**
- **Banks and lending institutions use it to forecast new homes purchases.**
- **Venture capital firms use it to forecast market potential/evaluate business plans.**
- **Meteorologists use it to predict precipitation, temperatures,** etc.

A comment on time scales:



- **Modern technology allows to record data on frequent time scales:**
  - ✓ **Stock data are available at ticker level.**
  - ✓ **Online and in store purchases are recorded in real time.**

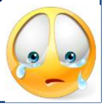
In considering choice of time scale, one must consider the scale of the required forecasts and the level of noise in data.

Question:
To forecast next-day sales at a grocery store, would you use minute-by-minute sales data or daily aggregates?

A comment on length of a time series:

- **Many modern time series are long:**
  - **Weekly interest rates,**
  - **Daily closing stock prices,**
  - **The electrical activity of the heart observed at millisecond intervals**

- **Other time series are short:**
  - **Annual data for FSU, starting 1991, about 30 observations;**
  - **Annual data for Israel, starting 1948, less than 100 observations;**
  - **Enrollments is PSTAT 115, very few observations …**



Your Example of a short and long time series?

PSTAT 174/274:
For meaningful time series analysis we want 100 or more
definitely not less than 50 observations;
Once we learn theory, we will know how to calculate the minimum number for our series.

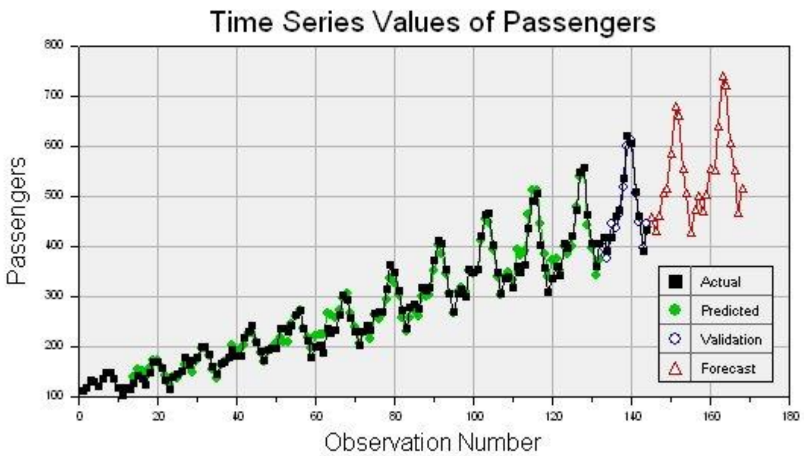
Time Series Data: Summary

REMEMBER: **The main difference between time series and other statistical samples:**

- ***dependent observations***

  that become available at

  - ***equally spaced time intervals*** **&**
    - ***are time-ordered***

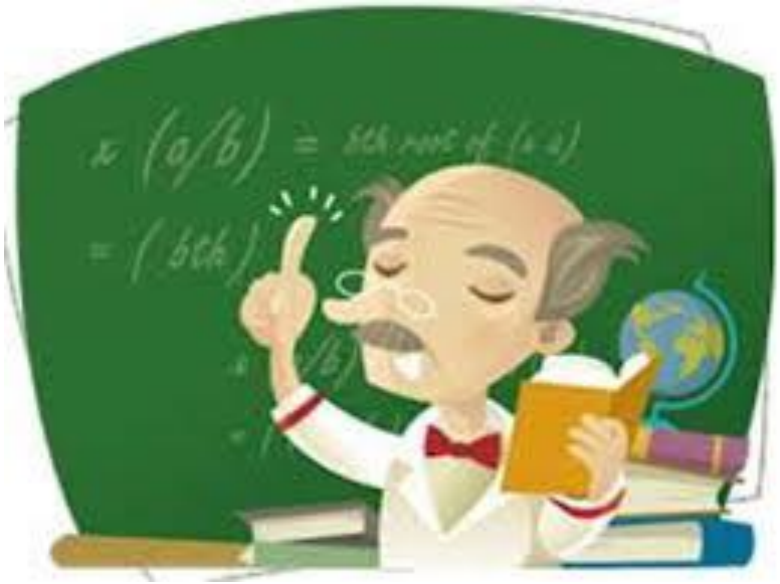


Goals of Time Series are *explanatory and predictive:*

- **understand or model stochastic mechanism that gives rise to an observed series**
- **forecast the future values of a series based on the history of that series**

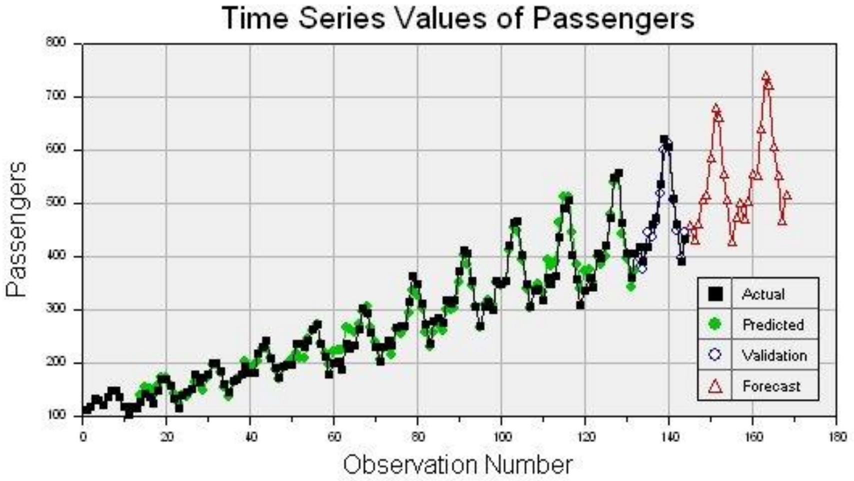
PSTAT 174/274: Intro to Time Series

Lecture 1, Part 3 : Trend and Seasonality

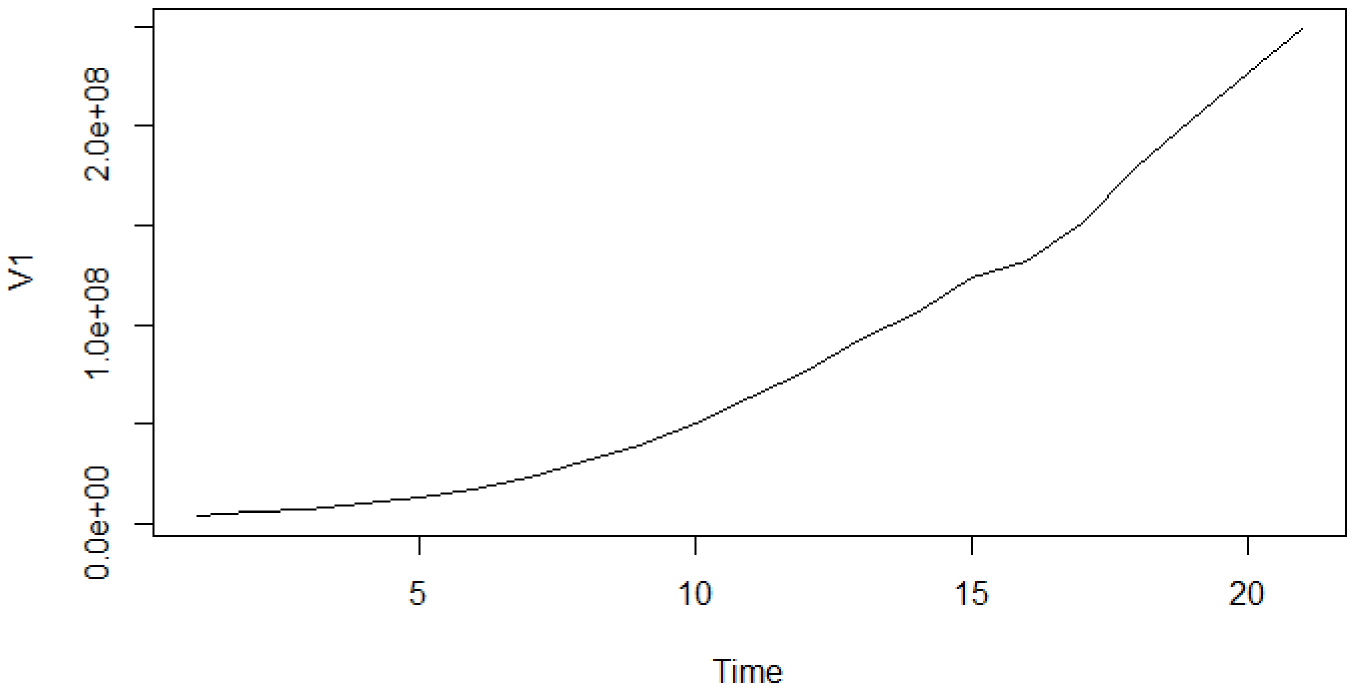


Part 3 of Lecture 1, Outline:

- **Examples of Time Series data**
- **Characteristics of Time Series Data:**
- **Trend and Seasonality**



US Population, 1790-1990, ten year intervals



X_t =population of US at year t: t=1790 X_t= 3,929,214

(in millions) t=1800 X_t= 5,308,483,

….

t=1990 X_t = 248,709,873

Note: looks like exponential or quadratic trend.

Non-stationary times series

US Population, 1790-1990, ten year intervals

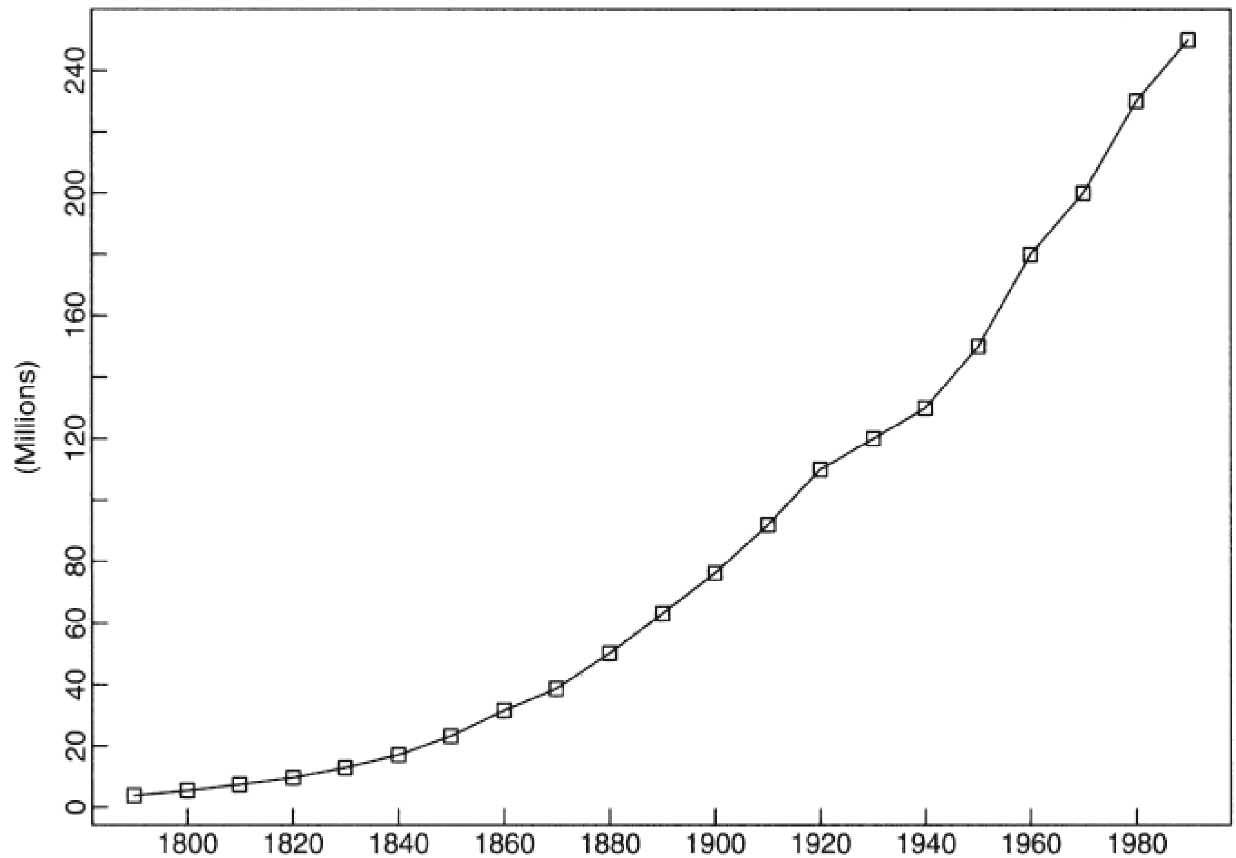
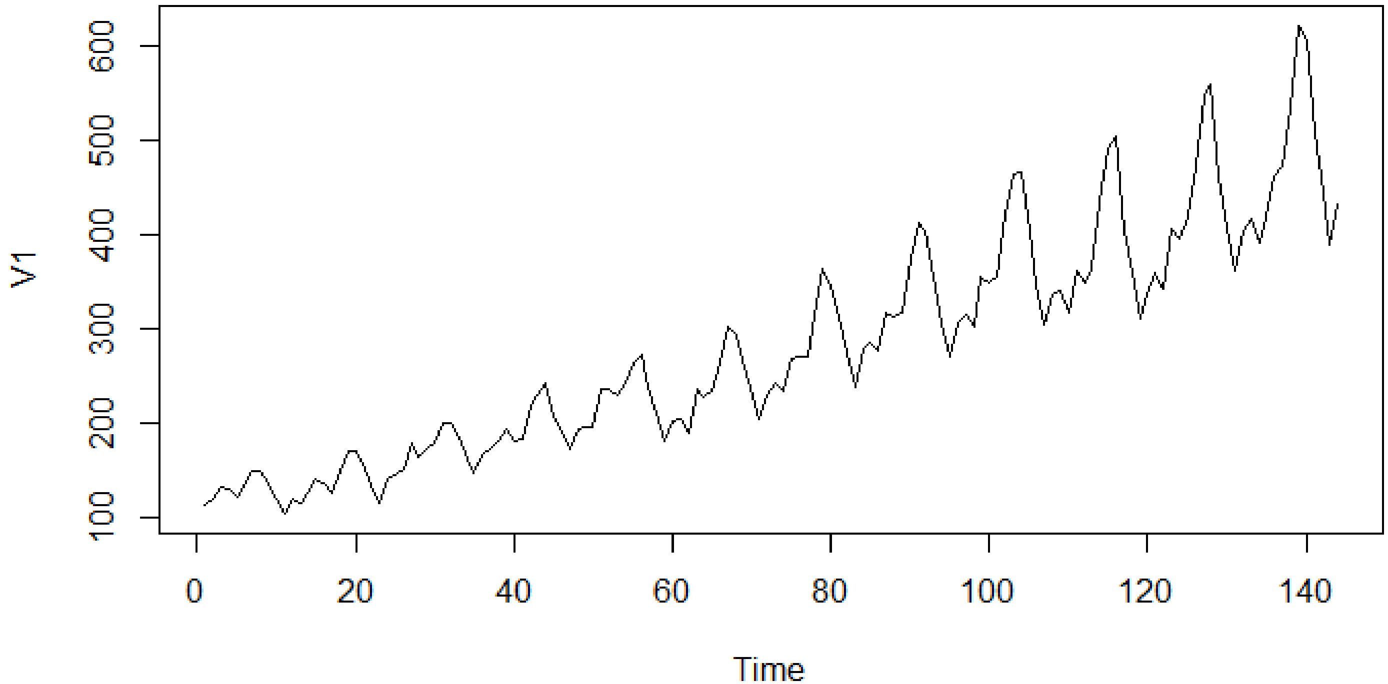


Figure 1-5
Population of the U.S.A. at ten-year intervals, 1790–1990.

(**taking a closer look:**
discrete points recording population of US every 10 years starting from 1790

International Airline Data.
Monthly totals of international passengers (1/1949 --12/1960)

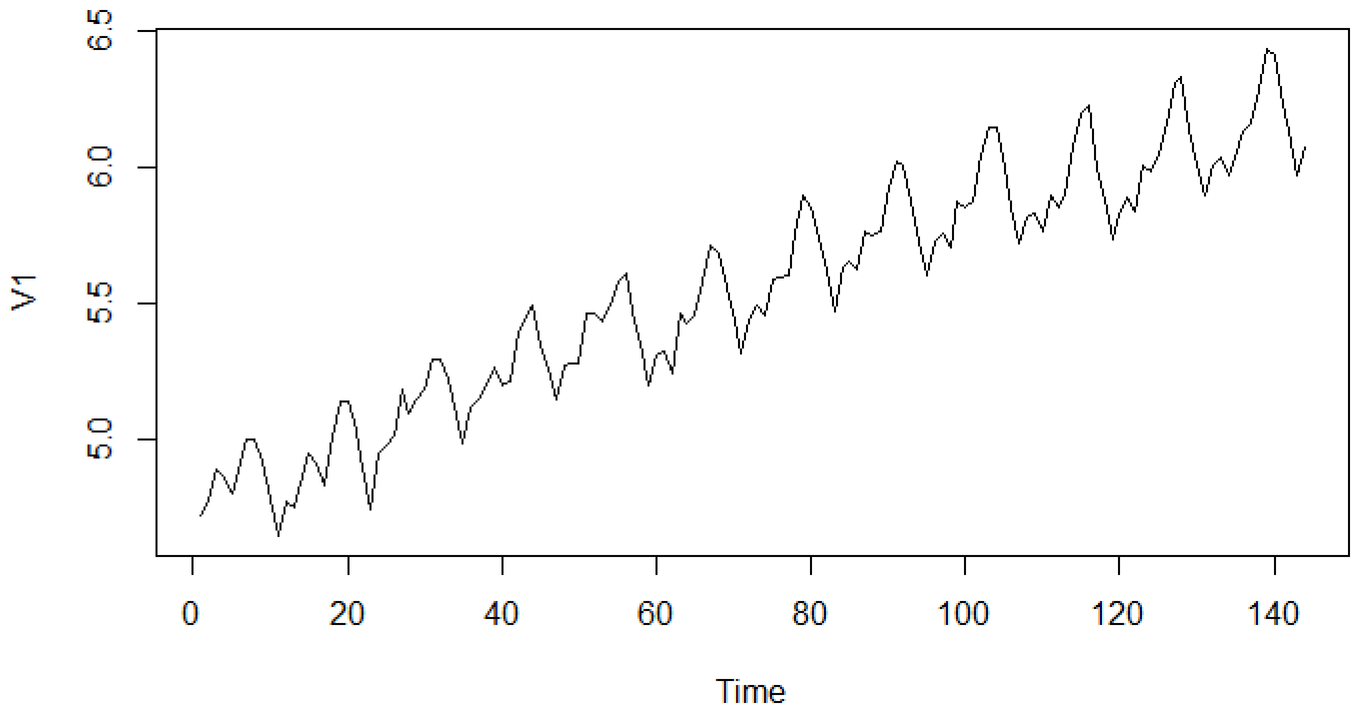


X_t=total number of passengers (in thousands) taking flights on the airline in a month t.

Non-stationary seasonal times series:

- **upward trend (linear?)**
- **Seasonal (low in winter, high in summer)**
- **Variability increases with time**

International Airline Data. Monthly totals of international passengers (1/1949 --12/1960)



X_t=total number of passengers (in thousands) taking flights on the airline in a month t.

V_t=log(X_t), t=1, 2, …, 144, natural logarithms of the AIRPASS.DAT file.

Trend and Seasonal Variation-summary

Trend: a systematic, non-periodic change in the time series.

- **Deterministic trends:** caused by non-random phenomenon,
  can be predicted with certainty,
  can be modeled by regression

  Linear trend: $X_t = a + bt + Z_t$

  Quadratic trend: $X_t = a + bt + ct^2 + Z_t$

- **Stochastic trends:** caused by random variation often induced by
  the dependence between adjacent variables.
  Analysis uses autocorrelations

Seasonal Variation: any recurrent pattern,
e.g., within each year in which the series is observed.

Error term:
variation in the time series not explained by the trend or seasonality.
Unlike in regression, the error terms are depended random variables.

Take a Break: Some Images

All slides of this class will use

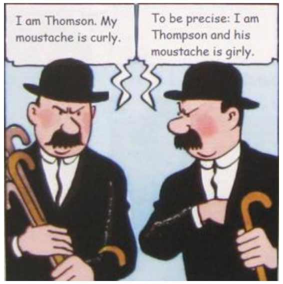
(call me silly, but I love comics. Tintin is my favorite together with Calvin and Hobbs and Garfield)

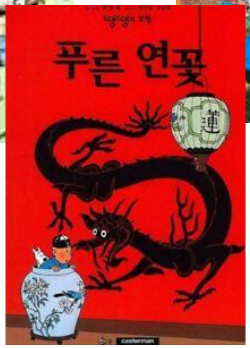
* **cartoon characters**  **for Check your Understanding slides and**  **for following answers**

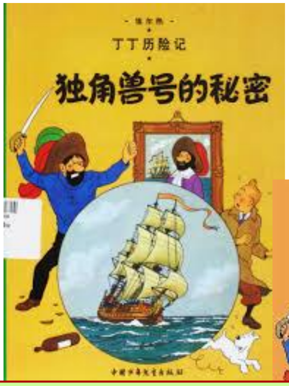
* **cartoon character named Professor Calculus**  **when reviewing math facts**

The Adventures of Tintin **are a series of comic books about boy-detective Tintin and his friends created by cartoonist Georges Remi, who wrote under the pen name Herge.**
https://en.wikipedia.org/wiki/The_Adventures_of_Tintin

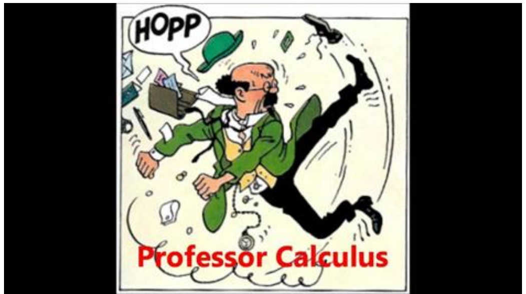
Tintin books are translated to most languages







This one was my first Tintin book and remains my favorite!



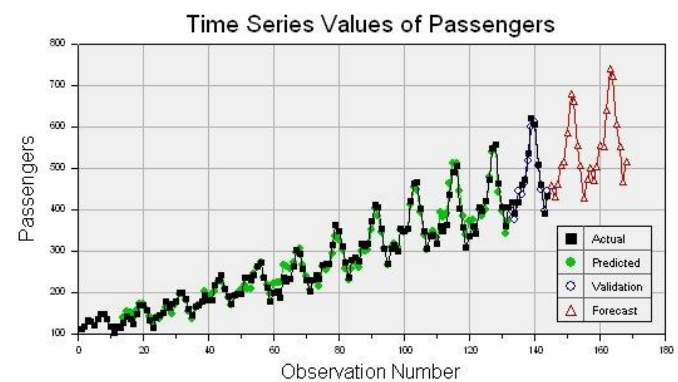
Trend and Seasonality Check of Understanding

Determine which of the following statements are true:

I. Deterministic trends can be modeled using regression.

II. Stochastic trends usually have a plausible physical explanation, such as an increase in population.

III. Short term extrapolation of deterministic trends can be justified by claiming that underlying trends will usually change slowly in comparison with the forecast lead time.



A. None are true

B. I and II only

C. I and III only

D. II and III only

E. The answer is not given by (A), (B), (C) or (D).

Check your answer of the next slide

Check your Understanding: Trend and Seasonality

Determine which of the following statements are true:

I. Deterministic trends can be modeled using regression.

II. Stochastic trends usually have a plausible physical explanation, such as an increase in population.

III. Short term extrapolation of deterministic trends can be justified by claiming that underlying trends will usually change slowly in comparison with the forecast lead time.

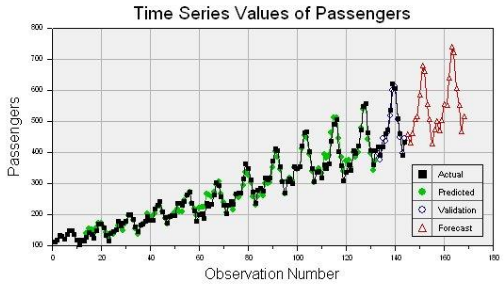
A. None are true

B. I and II only

C. I and III only

D. II and III only

E. The answer is not given by (A), (B), (C) or (D).



Short answer: C

Long answer:

I. is true: deterministic trend is non-random and can be modeled by regression;

III. is true: deterministic trend is a function estimated by regression; it changes slowly and can be extrapolated.

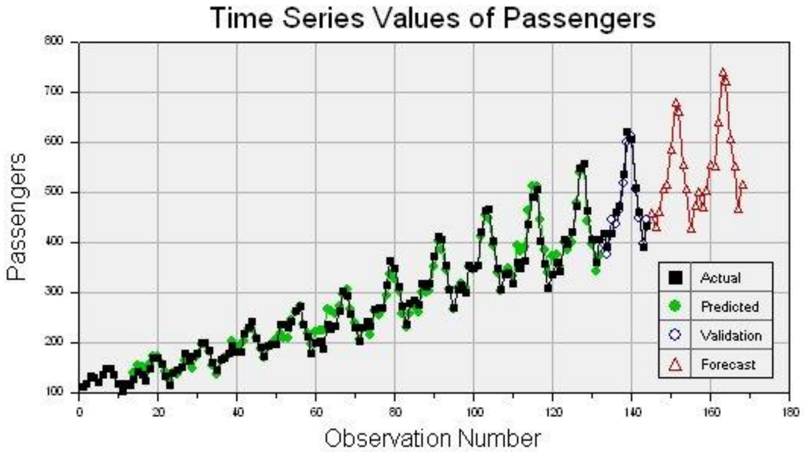
II. is incorrect: stochastic trend is caused by a random phenomena. The deterministic trends typically have plausible physical explanations.

REMEMBER: **The main difference between time series and other statistical samples:**

- *dependent observations*

  that become available at

  - *equally spaced time intervals* **&**
    - *are time-ordered*



Goals of Time Series are *explanatory and predictive:*

- **understand or model stochastic mechanism that gives rise to an observed series**
- **forecast the future values of a series based on the history of that series**

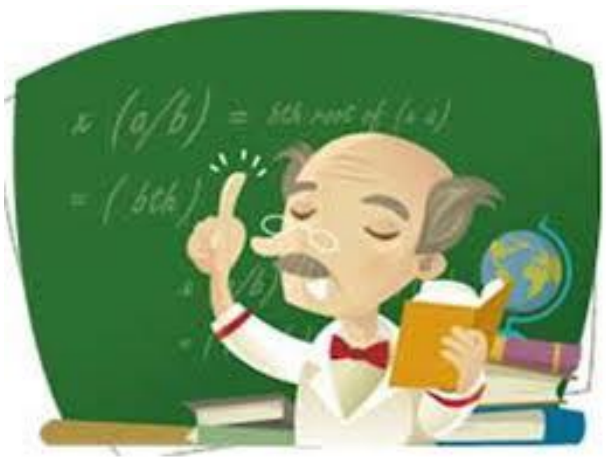
"Predicting the future is hard, especially if it hasn't happened yet." -- Yogi Berra

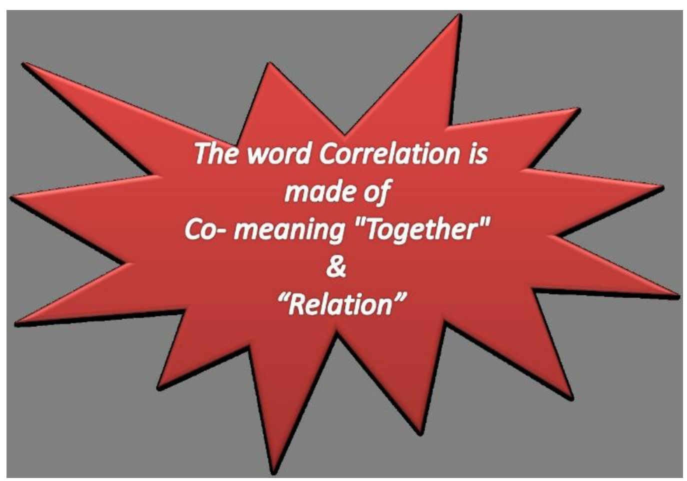
PSTAT 174/274: Intro to Time Series

Lecture 1, Part 4 : Autocovariance and Autocorrelation

Part 4 of Lecture 1, Outline:

- **Review of random vectors (X,Y)**
- **Covariance and Correlation**
- **Autocovariance and Autocorrelation**
- **White Noise**





Review: Characteristics of a Random Vector (X, Y)

(i) bivariate c.d.f. $F_{X,Y}(x, y) := P(X \leq x, Y \leq y))$

(ii) First and Second order moments:

- means $\mu_X = EX$, $\mu_Y = EY$
- variances $\sigma_X^2 = Var(X)$, $\sigma_Y^2 = Var(Y)$
- covariance and correlation:

$$\gamma(X, Y) = Cov(X, Y) \equiv Cov(Y, X) \stackrel{def}{=} E[(X - EX)(Y - EY)] \equiv E(XY) - E(X)E(Y).$$

$$\rho(X, Y) \equiv \rho_{XY} \equiv Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \equiv \frac{\gamma(X, Y)}{\sigma_X \sigma_Y}.$$

- If $Cov(X, Y) = 0$, then X and Y are uncorrelated.

  X and Y uncorrelated implies $\rho(X, Y) = 0$.
- $|\rho(X, Y)| \leq 1$
- Independent X and Y are always uncorrelated:

  $Cov(X, Y) = E[(X - EX)(Y - EY)] = E[(X - EX)] \cdot E[(Y - EY)] = 0$
- Uncorrelated X and Y might be dependent.

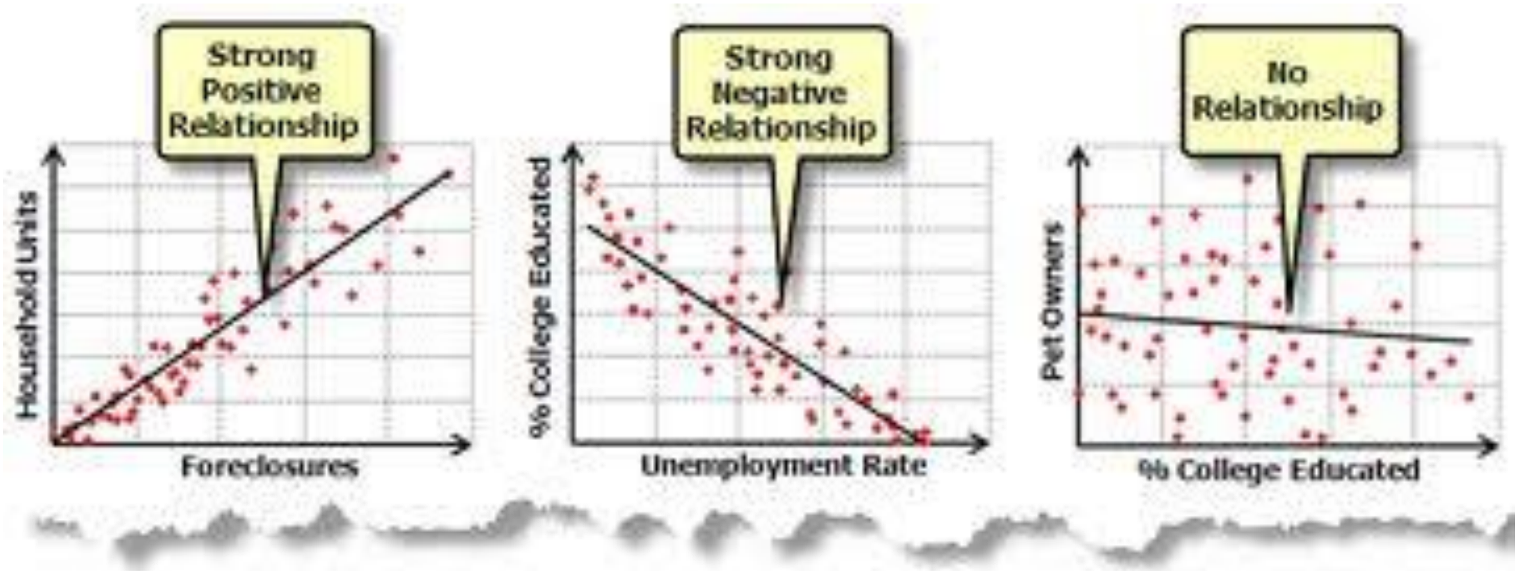
Meaning of
mean of X?
variance of X?

Example of dependent uncorrelated X and Y?

Importance of Correlation

Why?
See Lecture Notes

- For Uncorrelated X & Y, Var(X+Y) = Var(X) + Var(Y)
- If Cov(X,Y) ≠ 0, Var (X+Y) = Var(X) + Var(Y) + 2 Cov (X,Y)
- $\rho_{XY} \equiv$ Cor (X,Y) is a measure of linear dependence.
- Classical Regression and Time Series methods are based on ρ_{XY} :



Importance of Correlation: Estimating from Data

Estimate of the sample mean μ_X and variance σ^2_X from data $(x_1, \ldots, x_n)$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s^2_X \equiv \hat{\sigma}^2_X = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Estimate of the correlation coefficient ρ_{XY} from data $(x_1, y_1), \ldots, (x_n, y_n)$:

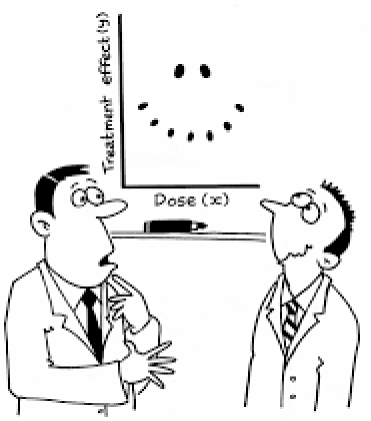
$$\hat{\rho}_{XY} = \left(\frac{1}{n-1}\right)\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{s^2_X s^2_Y}}.$$

Note: if observations in a sample are i.i.d., the order of summation is unimportant.

What to conclude from calculating $\rho_{XY} \equiv Cor(X,Y)$:

$|\rho_{XY}| \approx 1$ means strong linear relationship,
in a scatter diagram points (x_i, y_i) are close to a line;

$\rho_{XY} = 0$ means uncorrelated, no linear relationship
X & Y are either independent or
have nonlinear relationship

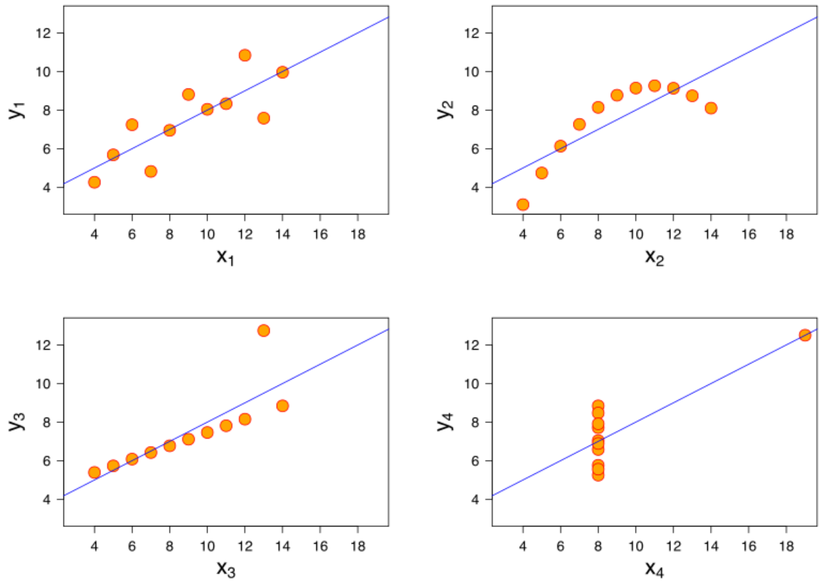


"It's a non-linear pattern with outliers.....but for some reason I'm very happy with the data."

Importance of Correlation

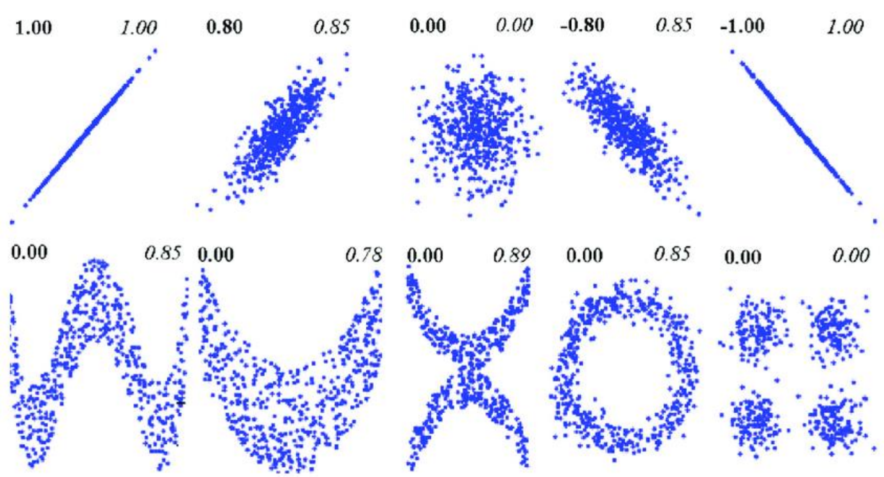
$\rho_{XY} = ?$

- $\rho_{XY} \equiv$ **Cor (X,Y) is a measure of linear dependence ONLY**



Four sets of data with the same correlation of 0.816, same mean & variance.

Anscombe, Francis J. (1973) Graphs in statistical analysis.
***American Statistician*, 27, 17–21.**



Vu, Tue, Mishra, A. & Kumar, G. (2018) Information Entropy Suggests Stronger Nonlinear Associations between Hydro-Meteorological Variables and ENSO.
***Entropy*, 20, 38. doi:10.3390/e20010038**

Moving to Time Series

- A series of values recorded in time

  $X_t, X_{t+h}, X_{t+2h}, X_{t+3h}, \ldots,$ or $X(t), X(t+h), X(t+2h), \ldots$

  h= time between observations or *sampling interval;*
- Order: very important! Observations dependent!
- Notation: position in the series: $X_1, X_2, X_3, \ldots, X_n, \ldots$
- Probabilistic/Statistical Description:

for each time index t = 1, 2, … define

- Mean $\mu_X(t) = E(X_t)$;
- Variance $\sigma_X^2(t) = Var(X_t)$;

Second-order properties

For each pair of time indices t and s find

- Autocovariance and Autocorrelation functions

ACVF: $\gamma_X(t,s) = Cov(X_t, X_s)$;

ACF: $\rho_X(t,s) = Cor(X_t, X_s)$.

Why 'auto'?

Review: Description of Time Series

- **Probabilistic/Statistical Description of time series $X_1, X_2, X_3, \ldots, X_n, \ldots$**

To describe a time series $\{X_t, t = 1, 2, \ldots\}$ we define

(i) **The finite-dimensional distributions**: (fi-di) d.f. is the joint d.f. for the vector $(X_{t_1}, \ldots, X_{t_n})$:

$$F_{t_1 \ldots t_n}(x_1, \ldots, x_n) = P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n), \forall n \ge 1, \forall t_1 < t_2 < \ldots < t_n.$$

(ii) **First- and Second-order moments.**

- Mean: $\mu_X(t) = EX_t$
- **Variance:** $\sigma_X^2(t) = E(X_t - \mu_X(t))^2 \equiv EX_t^2 - \mu_X(t)^2$
- **Autocovariance function (acvf):**

$$\gamma_X(t, s) = Cov(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))] \equiv E(X_t X_s) - \mu_X(t)\mu_X(s)$$

(Note: this is an infinite matrix).

- **Autocorrelation function (acf):**

$$\rho_X(t, s) = Cor(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} = \frac{\gamma_X(t, s)}{\sigma_X(t)\sigma_X(s)}$$

Properties of the process which are determined by the first- and second- order moments are called second-order properties.

Example 1.3.1: White Noise WN(0, σ^2)

Notation: $\{Z_t\}$

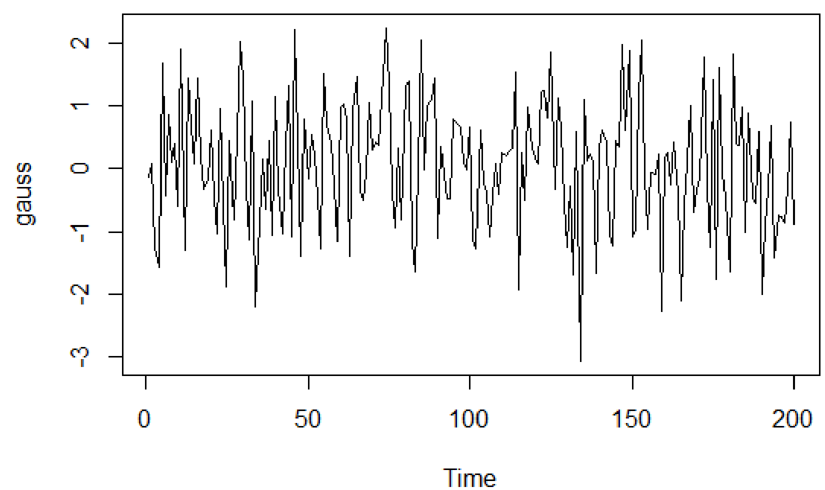
Definition of WN: Z_t are mean zero, constant variance and uncorrelated:

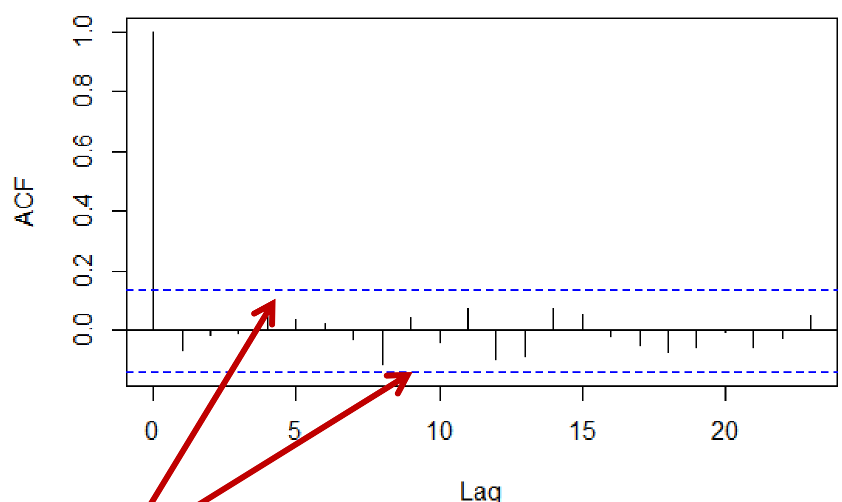
$\mu_Z(t) = E(Z_t) = 0$, $\gamma_Z(t,s) = Cov(Z_t, Z_s) = 0$ if $t \neq s$ (uncorrelated);

$\gamma_Z(t,t) = Cov(Z_t, Z_t) = \sigma_Z^2 = Var(Z_t)$;

Gaussian WN: $Z_t \sim N(0, \sigma_Z^2)$.

Are Gaussian WN entries IID?
If yes, why?
Are non-Gaussian WN entries IID?





No trend or seasonality. Very choppy.
ACF = 0 for lags $k = |t-s| > 0$.

95% confidence interval based on Normal distribution.

What to remember from Lecture 1:

- **Time series observations are time-ordered, equally (in time) spaces and dependent.**
- **The main goal of Time Series analysis is forecasting.**
- **Visual characteristics of non-stationary data:**
  - **Trend: a systematic, non-periodic change in the time series.**
  - **Seasonality: any recurrent pattern**
  - **Variance change with time**
- **Review meaning of mean, variance, covariance and correlation for (X,Y)**
- **Notions of autocovariance and autocorrelation (slides 40-41)**
- **The first Time Series model: White Noise (WN)   (slide 42)**

Welcome to Lecture 2

Outline of Lecture 2

Part I	***Review of Lecture 1:***	***pp. 45 – 47***
	Example 1.3.2: Smoothing of White Noise	***p. 48***
Part II	***Stationary Time Series***	***pp. 50, 52-54***
	Review of Bivariate Normal Distribution	***p. 51***
	Random Walk	***p. 55***
	Check your understanding slides	***p. 56 -- 58***
Part III	***MA models:***	***pp. 60-66***
	MA(1) model, definition and ACF	***pp. 61, 64***
Main concepts of lecture 2:		*p. 67*
R code:		*p. 68*

What are the most important concepts of Lecture 1?

What are the most important concepts of Lecture 1?

Hints:

- **What is time series?**
  - ▪ order of observations ▪ in/dependence of observations

- **Visible characteristics of time series**
  - ▪ length ▪ trend ▪ seasonality ▪ scale

- **Second-order probabilistic descriptions**
  - ▪ mean ▪ variance ▪ autocovariance ▪ autocorrelation

- **Example 1.3.1: WN and Gaussian WN**

Review of Example 1.3.1: White Noise WN(0, σ^2)

Notation: $\{Z_t\} \sim WN(0, \sigma_Z^2)$

Definition: **Random variables Z_t, t=1, 2, …, are**

- **mean zero: $\mu_Z(t) = E(Z_t) = 0$;**
- **constant variance: $Var(Z_t) = \sigma_Z^2$**

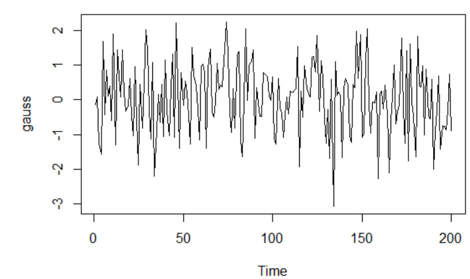
  (Variance does not change with time t; Variance is the same for all Z_t)
- **uncorrelated: $\gamma_Z(t,s) = Cov(Z_t, Z_s) = 0$ if $t \neq s$ (uncorrelated).**

  (Note: when s=t, $\gamma_Z(t,t) = Cov(Z_t, Z_t) = Var(Z_t) = \sigma_Z^2$)

Special Case -- Gaussian WN $\{Z_t\}$:

- **It is WN, that is all three properties listed above hold;**
- **In addition, we specify distribution of Z_t : $Z_t \sim N(0, \sigma_Z^2)$ (normal).**

Are Gaussian WN entries IID? If yes, why?
Are non-Gaussian WN entries IID?



Review of Example 1.3.1: White Noise WN(0, σ^2)

Definition: $\{Z_t\} \sim WN(0, \sigma_Z^2)$ if the following 3 properties hold:

- **mean zero**: $\mu_Z(t) = E(Z_t) = 0$;
- **constant variance:** $Var(Z_t) = \sigma_Z^2$ (Variance is the same for all Z_t)
- **uncorrelated:** $\gamma_Z(t,s) = Cov(Z_t, Z_s) = 0$ if $t \neq s$ (uncorrelated).

Special Case -- Gaussian WN $\{Z_t\}$:

- $\{Z_t\} \sim WN(0, \sigma^2)$ and in addition, $Z_t \sim N(0, \sigma_Z^2)$ (normal) for each t.

What makes Gaussian WN special?

- **Fact to remember:** If Z_t and Z_s $(t \neq s)$ are uncorrelated, that is, $\gamma_Z(t,s) = Cov(Z_t, Z_s) = 0$, they do not have linear dependence but might still have nonlinear dependence;
- **Fact to remember:** If Z_t and Z_s are Normal (Gaussian) and are uncorrelated, they are independent. This is a property of Normal (Gaussian) distribution.

Important for calculations: For $\{Z_t\} \sim WN(0, \sigma^2)$,

$\gamma_Z(t,s) = Cov(Z_t, Z_s) = E(Z_tZ_s) - E(Z_t)E(Z_s) = E(Z_tZ_s) - 0 = 0$ if $t \neq s$. That is, $E(Z_tZ_s) = 0, t \neq s$.

$Var(Z_t) = E(Z_t^2) - [E(Z_t)]^2 = E(Z_t^2) - 0 = E(Z_t^2) = \sigma_Z^2$. That is, $E(Z_t^2) = \sigma_Z^2$ for all t

Example 1.3.2: Smoothing of WN(0, σ^2)

Goal: Calculate AVCF of the smoothed Gaussian WN of Example 1.3.1:

$$X_t = (1/3)(Z_{t-1} + Z_t + Z_{t+1})$$

$E(X_t) = (1/3)(E(Z_{t-1}) + E(Z_t) + E(Z_{t+1})) = 0$

$\gamma_X(t,s) = \mathrm{Cov}(X_t, X_s) = E(X_t X_s)$

WHY?

$= (1/9)\, E[(Z_{t-1} + Z_t + Z_{t+1})(Z_{s-1} + Z_s + Z_{s+1})]$

$= (1/9)\, E[(Z_{t-1} Z_{s-1} + Z_t Z_s + Z_{t+1} Z_{s+1})$
$+ (Z_{t-1} Z_s + Z_t Z_{s+1}) + (Z_t Z_{s-1} + Z_{t+1} Z_s)$
$+ (Z_{t-1} Z_{s+1} + Z_{t+1} Z_{s-1})]$

For $\{Z_t\} \sim WN(0, \sigma^2)$,
$E(Z_t) = 0$ for all t
$E(Z_t Z_s) = 0$, $t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

$$\gamma_X(t,s) = \begin{cases} 3/9\sigma_Z^2, & \text{if } t = s \text{ that is, } |t-s| = 0; \\ 2/9\sigma_Z^2, & \text{if } t-1 = s \text{ or } t = s-1, \text{ that is, } |t-s| = 1; \\ 1/9\sigma_Z^2, & \text{if } t-1 = s+1 \text{ or } t+1 = s-1, \text{ that is, } |t-s| = 2; \\ 0, & \text{if } |t-s| > 2. \end{cases}$$

Welcome to Lecture 2, Part 2

Outline of Lecture 2, Part 2:

- **Strict and (Weak) Stationarity, p. 50**
- **Review of Bivariate Normal Distribution, p. 51**
- **Stationary Time Series, pp. 52 – 56**
- **Check your understanding, pp. 56 -- 58**

From https://www.dictionary.com/browse/stationary :

Stationary **[stey-shuh-ner-ee]**

adjective

- ***standing still; not moving.***
- ***having a fixed position; not movable.***
- ***established in one place; not itinerant or migratory.***
- ***remaining in the same condition or state; not changing: The market price has remained stationary for a week.***
- ***geostationary.***

Strict Stationarity and (Second-order , Weak) Stationarity

Strict Stationary:

For any $n \geq 1$ and any set of times $t_1 < t_2 < \ldots < t_n$,

$$(\mathrm{X}_{\mathrm{t}_1}, \ldots, \mathrm{X}_{\mathrm{t}_n}) \stackrel{\mathcal{D}}{=} (\mathrm{X}_{\mathrm{t}_1+\mathrm{k}}, \ldots, \mathrm{X}_{\mathrm{t}_n+\mathrm{k}}) \text{ for all } \mathrm{k} \geq 1,$$

that is, the distribution does not change with time translation.

Weak stationarity:

(a) $E|X_t|^2 < \infty$; (b) $EX_t = \mu$ for all $t \in T$
(c) $\gamma_X(t, s) = \gamma_X(t + r, s + r) = \gamma_X(t - s)$ for all $t, s, r \in T$.

- strict stationarity + finite second moment imply stationarity
- stationarity does not imply strict stationarity

Fact: For Gaussian TS *stationarity* = *strict stationarity*.

WHY?

(b/c Gaussian distribution is determined by its mean and covariance)

Let's review some facts on Gaussian distribution – next slide

Review: Bivariate Normal Distribution

2.2 Facts about bivariate normal distribution.

(i) Continuous r.v.'s X and Y have a bivariate normal distribution with the parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ $(-1 \le \rho \le 1, \sigma_i^2 \ge 0)$ if their joint p.d.f. $f(x, y)$ is

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[(\frac{\mathrm{x}-\mu_1}{\sigma_1})^2 - 2\rho(\frac{\mathrm{x}-\mu_1}{\sigma_1})(\frac{\mathrm{y}-\mu_2}{\sigma_2}) + (\frac{\mathrm{y}-\mu_2}{\sigma_2})^2]\}$$

(ii) Facts:

(a) the marginal distributions of X and Y are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively;

(b) the $\mathrm{Cor}(X, Y) = \rho$.

(c) X and Y are independent iff $\rho = 0$.

(d) the conditional distribution of Y given $X = x$, is

$$N(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), (1-\rho^2)\sigma_2^2)$$

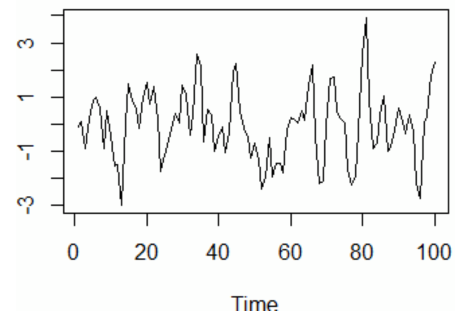
(e) the conditional distribution of X given $Y = y$, is

$$N(\mu_1 + \rho(\sigma_1/\sigma_2)(y - \mu_2), (1-\rho^2)\sigma_1^2)$$

Stationary Time Series

- **No trend**: Mean $\mu_X(t) = E(X_t) = \mu_X$ (constant)
- **No change of variance**: Variance $\sigma_X^2(t) = Var(X_t) = \sigma_X^2$ (constant)
- **No seasonality, no sharp change of behavior**
- **ACVF is a function of a lag:**

  $$\gamma_X(t, s) = Cov(X_t, X_s) = \gamma_X(t - s) = \gamma_X(k) \text{ with } |t - s| = k:$$



- **ACF is a function of a lag:**

  $$\rho_X(t, s) = Cor(X_t, X_s) = \rho_X(k) \text{ with } |t - s| = k.$$

- **Some Formulas to Remember:**

$$\gamma_X(t, s) = Cov(X_t, X_s) = E(X_t X_s) - E(X_t) E(X_s) = Cov(X_s, X_t) = \gamma_X(s, t) ;$$

$$\gamma_X(k) = \gamma_X(-k); \qquad \gamma_X(0) = Cov(X_t, X_t) = Var(X_t) = \sigma_X^2 ;$$

$$\rho_X(k) = Cov(X_t, X_s) / [\sigma_X(t)\, \sigma_X(s)] = \gamma_X(k) / \gamma_X(0);$$

$$\rho_X(k) = \rho_X(-k) ; \qquad \rho_X(0) = 1.$$

Review: Strict and (Weak) Stationarity

Review of two examples:

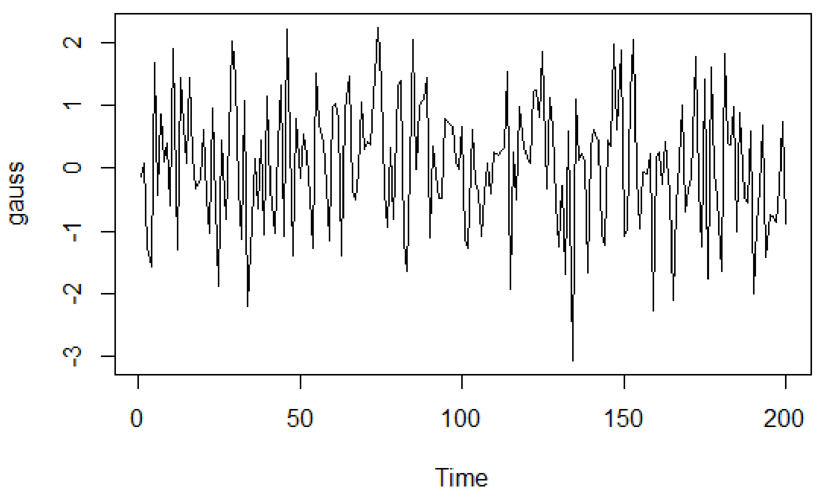
Example 1.3. : White Noise Z_t ~ WN $(0, \sigma_Z^2)$ is a TS such that

$E(Z_t) = 0$, $\gamma(s, t) = 0$ if $s \neq t$ and $\gamma(t, t) = Var(Z_t) = \sigma_Z^2$ if $s = t$

Question: is it stationary? Strictly stationary?

If White Noise Z_t is Gaussian WN if $Z_t \sim N(0, \sigma_Z^2)$ for all t.

Question: is it stationary? Strictly stationary?



(Weakly) Stationary TS:
$E(X_t) = \mu_X$ (constant)
$\gamma_X(t, s) = Cov(X_t, X_s) = \gamma_X(t - s) = \gamma_X(k)$, $|t - s| = k$;
$\sigma_X^2(t) = Var(X_t) = \gamma_X(0) = \sigma_X^2$ (constant)

Does knowing that Z has Gaussian distribution make a difference?

Review: Strict and (Weak) Stationarity

Example 2.1.4: **Smoothed Gaussian WN $X_t = (1/3)(Z_{t-1} + Z_t + Z_{t+1})$**

In 1.3.2, we calculated ACVF:

$$\gamma_X(t,s) = \begin{cases} 3/9\,\sigma_Z^2 = 1/3\sigma_Z^2, \text{ if } t = s \text{ that is, } |t-s| = 0; \\ 2/9\sigma_Z^2, \text{ if } t-1 = s \text{ or } t = s-1, \text{ that is, } |t-s| = 1; \\ 1/9\sigma_Z^2, \text{ if } t-1 = s+1 \text{ or } t+1 = s-1, \text{ that is, } |t-s| = 2; \\ 0, \quad \text{if } |t-s| > 2. \end{cases}$$

Question: is this time series stationary? Strictly stationary?

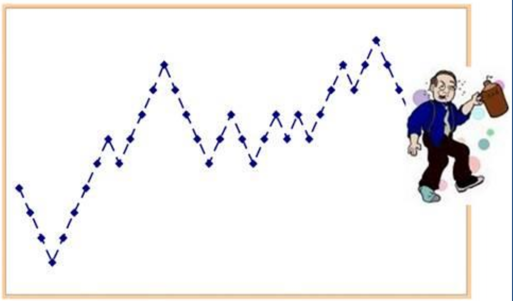
Calculate ACF for this series

For $\{Z_t\} \sim WN(0, \sigma^2)$,
$E(Z_t) = 0$ for all t
$E(Z_t Z_s) = 0,\ t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

(Weakly) Stationary TS:
$E(X_t) = \mu_X$ (constant)
$\gamma_X(t, s) = Cov(X_t, X_s) = \gamma_X(t-s) = \gamma_X(k),\ |t-s| = k$;
$\sigma_X^2(t) = Var(X_t) = \gamma_X(0) = \sigma_X^2$ (constant)

Review: Strict and (Weak) Stationarity

Example 2.1.5: Let $Z_t \sim$ IID $(0, \sigma_Z^2)$ be I.I.D. WN, that is, Z_t are i.i.d. r.v.s with mean zero and variance σ^2.



Random Walk (RW) is defined as $X_t = Z_1 + \ldots + Z_t$.

Calculate:

- $E(X_t) = 0$,
- Take $s = t + k$, $k \geq 0$. Then

$$\gamma_X(t, s) = E(X_t X_s) = E[(Z_1 + \ldots + Z_t)(Z_1 + \ldots + Z_t + Z_{t+1} + \ldots + Z_{t+k})]$$
$$= E[(Z_1 + \ldots + Z_t)^2] + E[(Z_1 + \ldots + Z_t)(Z_{t+1} + \ldots + Z_{t+k})]$$
$$= \mathrm{Var}(Z_1 + \ldots + Z_t) + 0 = t\,\sigma^2 .$$

- Summary:
  - $\gamma_X(t, s) = \sigma^2 \min(t,s)$ or $\sigma^2 t$ for $s = t+k$, $k \geq 0$.
  - In particular, $\sigma_X^2 = \gamma_X(t, t) = \sigma^2 t$.

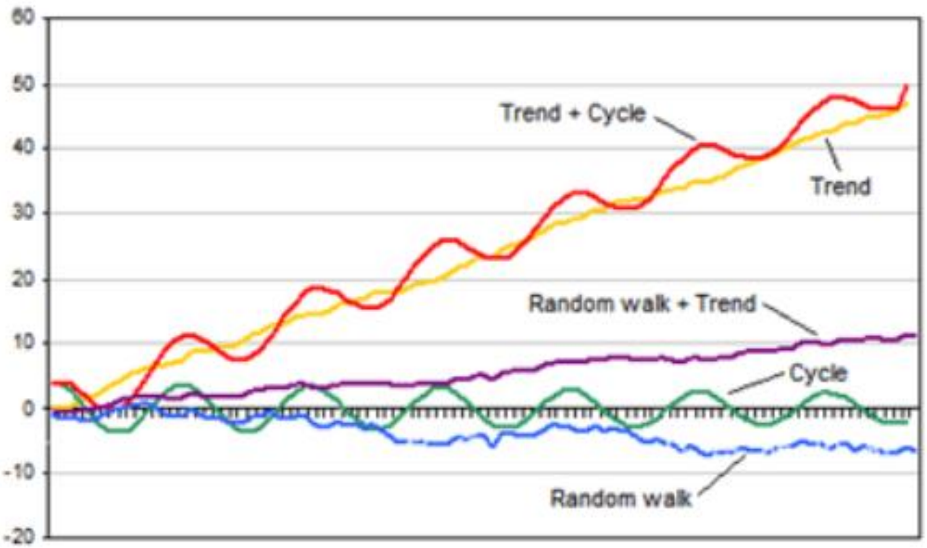
For $\{Z_t\} \sim WN(0, \sigma^2)$,
$E(Z_t) = 0$ for all t
$E(Z_t Z_s) = 0$, $t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

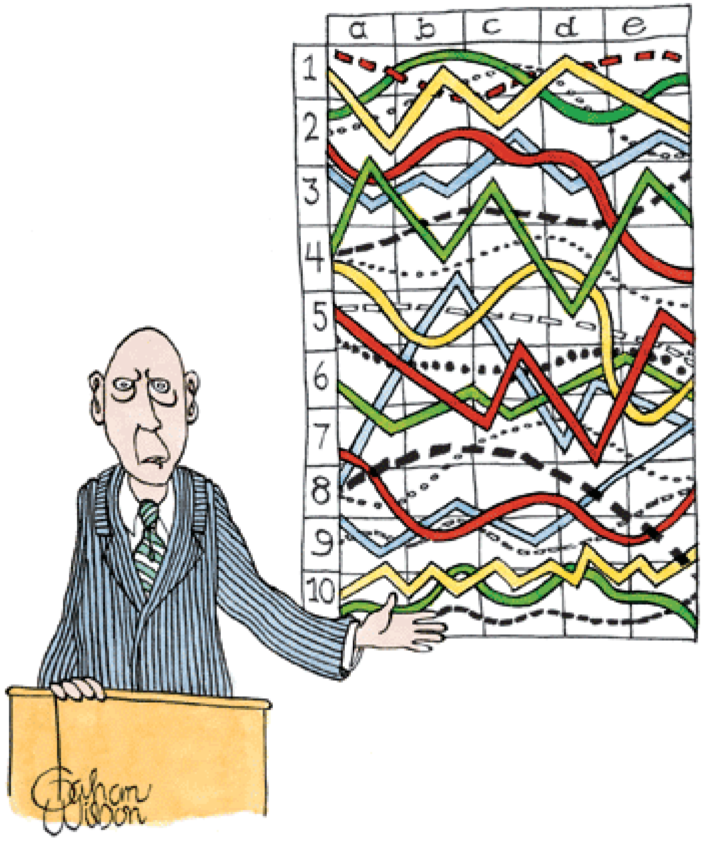
Question: is RW stationary? Strictly stationary?

Stationary Time Series

What are the visual characteristics of a stationary time series?

Some hints:





"I'll pause for a moment so you can let this information sink in."

Stationary? Gaussian? Check your Understanding

Disclaimer: *these questions are meant to confuse you, read them carefully and THINK! Did you get all details right? Remember definitions? Paid attention?*

True or False:

I. **All random walk processes are nonstationary.**
II. **All random walk processes are stationary. (Meaning weakly stationary.)**
III. **All random walk processes are strictly stationary.**
IV. **For a stationary process its mean does not depend on time, i.e., a constant.**
V. **For a stationary process variance is a constant but can be infinite.**
VI. **A strictly stationary process is always weakly stationary.**
VII. **A strictly stationary Gaussian process is always weakly stationary.**
VIII. **A weakly stationary Gaussian process is always strictly stationary.**
IX. **A seasonal time series may be stationary.**
X. **A time series with trend may be stationary.**
XI. **A Gaussian random walk is stationary.**

Check your answer of the next slide

Stationary? Gaussian? Check your Understanding

True or False:

I. All random walk processes are nonstationary. True

II. All random walk processes are stationary. (Meaning weakly stationary.) False

III. All random walk processes are strictly stationary. False

IV. For a stationary process its mean does not depend on time, i.e., a constant. True

V. For a stationary process variance is a constant but can be infinite. False

VI. A strictly stationary process is always weakly stationary. False

VII. Any weakly nonstationary process is automatically strictly nonstationary. False

VIII. A strictly stationary Gaussian process is always weakly stationary. True

IX. A weakly stationary Gaussian process is always strictly stationary. True

X. A seasonal time series may be stationary. False

XI. A time series with trend may be stationary. False

Explanation for I-III: We showed in Example 2.1.5 that the variance of the RW depends on t, that is, not a constant: $\sigma_X^2\,(t) = \sigma_Z^2\,t$. Thus, it is not weakly or strictly stationary.

For IV-VII: Weakly stationary processes must have constant and finite mean and variance.

VI is False b/c strictly stationary processes might have ∞ variance, e.g., i.i.d. Cauchy r.v.s.

VII is False in cases when time series is strictly stationary but has infinite variance.

For VIII-IX: The Gaussian distribution is fully determined by its finite mean and correlation.

For X-XI: Time series with trend/seasonality have non constant means, i.e., non-stationary.

Welcome to Lecture 2, Part 3

Outline of Lecture 2, Part 3:

- **Moving Average Process MA(0), p. 60**
- **Moving Average Process MA(1), pp. 61 -- 66**
- **Main concepts of Lecture 2, p. 67**
- **R code, p. 68**

Moving Average? Moving where?

Moving Average Models: MA(0)

https://hbr.org/2020/03/understanding-the-economic-shock-of-coronavirus
Harvard Business Review, 3/27/2020

<u>3.1</u>: White Noise (also called "shocks") AGAIN?

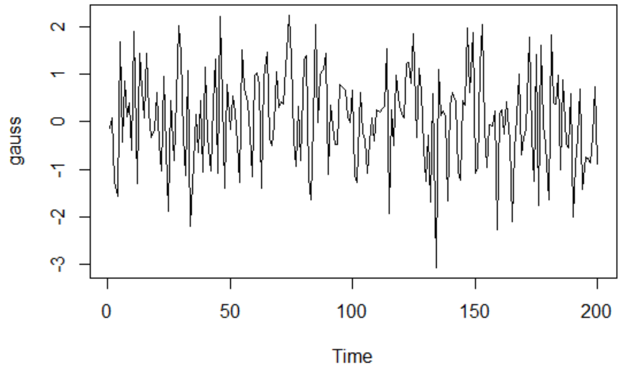
It must be important …

$Z_t \sim WN(0, \sigma_Z^2)$ is a TS such that

$$E(Z_t) = 0, \gamma(s, t) = E(Z_t Z_s) = \sigma_Z^2 \text{ if } s = t \text{ \& } \gamma(s, t) = 0 \text{ if } s \neq t$$

$Z_t \sim$ IID $(0, \sigma_Z^2)$ is a WN with independent observations;

If White Noise Z_t is Gaussian WN if $Z_t \sim N(0, \sigma_Z^2)$ for all t.



WN is also a MA(0) model.
Read: Moving Average of Order 0

Moving Average Models: MA(1)

3.2: MA(1) Moving Average of Order One

$$X_t = Z_t + \theta_1 Z_{t-1}$$

where $Z_t \sim WN(0, \sigma_Z^2)$ and $|\theta_1| < 1$.

In MA(1) model, current observation X_t depends on today's shock Z_t and yesterday's shock Z_{t-1} !

MA(1) processes are created from WN components!

Examples of MA(1):

$$X_t = Z_t + 0.8\, Z_{t-1}$$
$$X_t = Z_t - 0.6 Z_{t-1}$$
$$X_t = Z_t + 0.2 Z_{t-1}$$

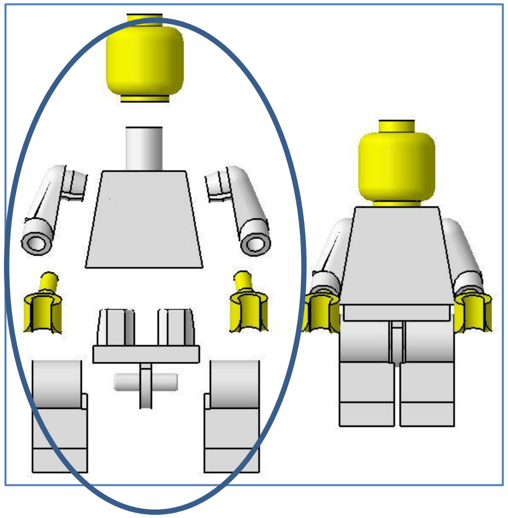
Can you give an intuitive reason why θ_1 should be in the (-1,1) interval ?

Creating complicated from basic: MA(1) from WN

Lego blocks **White Noise** Z_t, $t=1,2,\ldots$

From Lego blocks construct
different Lego people

From White Noise construct
different TS models



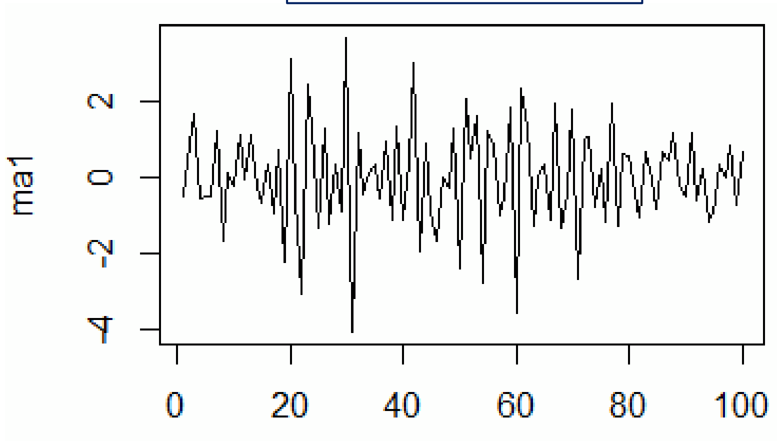
Z_t, Z_{t-1} $X_t = Z_t + 0.95\, Z_{t-1}$

Boy

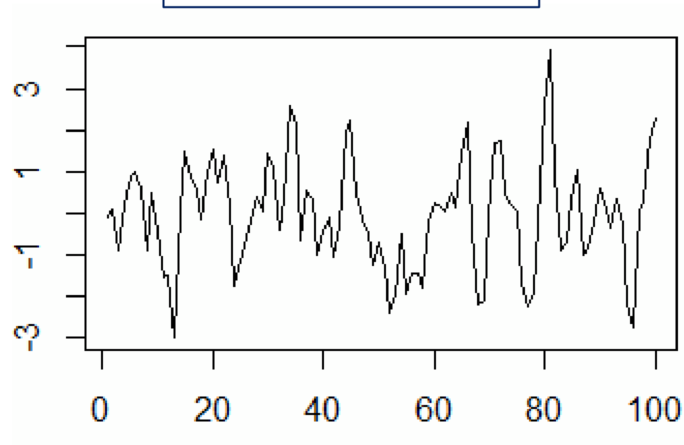
MA(1)

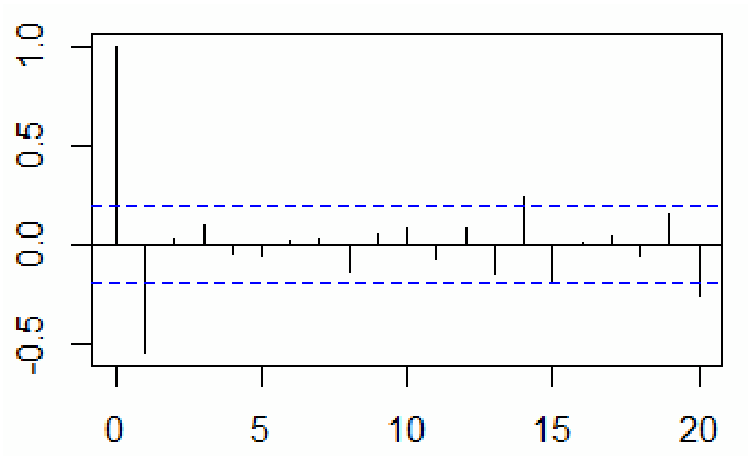
100 simulated values of MA(1) and its ACF

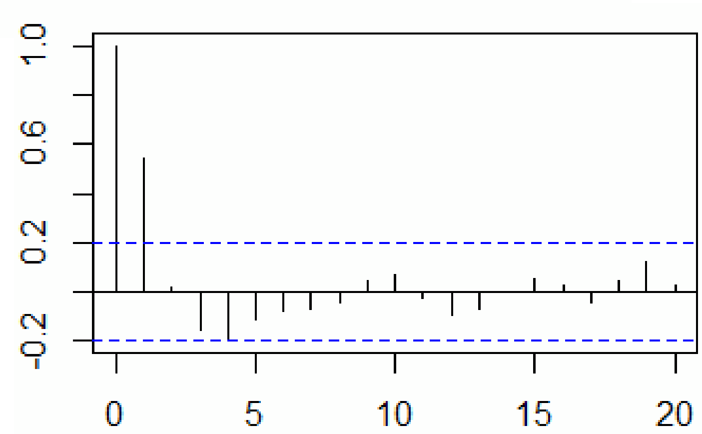
$X_t = Z_t - 0.95\, Z_{t-1}$



$X_t = Z_t + 0.95\, Z_{t-1}$







$X_t = Z_t - 0.95\, Z_{t\text{-}1}$
Correlation is 0 after lag 1.
Graph rough -- strong negative correlation

$X_t = Z_t + 0.95\, Z_{t-1}$
Correlation is 0 after lag 1.
Graph smooth -- strong positive correlation

R Commands: summarized to the end

ACVF and ACF of Moving Average Models: MA(1)

3.2: MA(1) Moving Average of Order One

$X_t = Z_t + \theta_1 Z_{t-1}$ where Z_t ~ WN $(0, \sigma_Z^2)$ and $|\theta_1| < 1$.

Calculate second moments:

$EX_t = E(Z_t + \theta_1 Z_{t-1})$ = 0 b/c expectation is linear and E Z_t = 0

$\gamma_X(k) = E\{(Z_t + \theta_1 Z_{t-1})\,(Z_{t+k} + \theta_1 Z_{t+k-1})\}$

$= E\{(Z_t Z_{t+k} + \theta_1{}^2 Z_{t-1} Z_{t+k-1})$

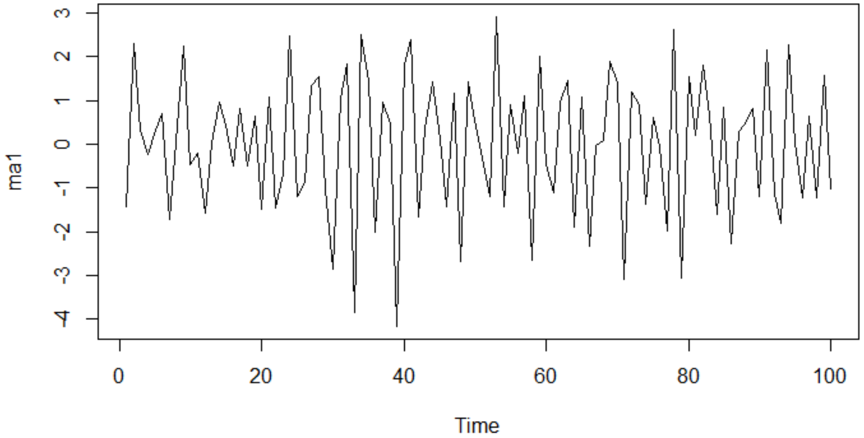
$+ \theta_1 (Z_t Z_{t+k-1} + Z_{t-1} Z_{t+k})\}$

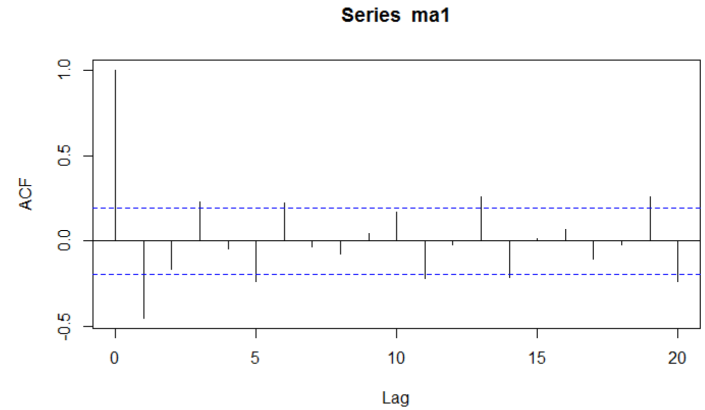
$$= \sigma_Z^2 \begin{cases} (1+\theta_1^2) & \text{if } k = 0; \\ \theta_1 & \text{if } k = -1 \text{ or } 1, \text{ that is, } |k| = 1; \\ 0 & \text{if } |k| > 1. \end{cases}$$

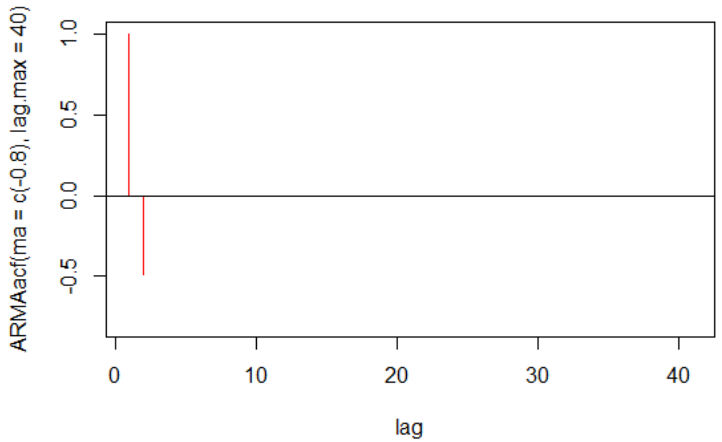
For $\{Z_t\} \sim$WN$(0, \sigma^2)$,
$E(Z_t) = 0$ for all t
$E(Z_t Z_s) = 0$, $t \neq s$.
$E(Z_t^2) = \sigma_Z^2$ for all t

X is stationary with $\mu_x = 0$, and ACF $\rho_X(0) = 1$; $\rho_X(k) = 0$ for $|k| > 1$, and $\rho_X(1) = \theta_1/(1+\theta_1{}^2)$

100 simulated values of MA(1) $X_t = Z_t - 0.8\, Z_{t-1}$





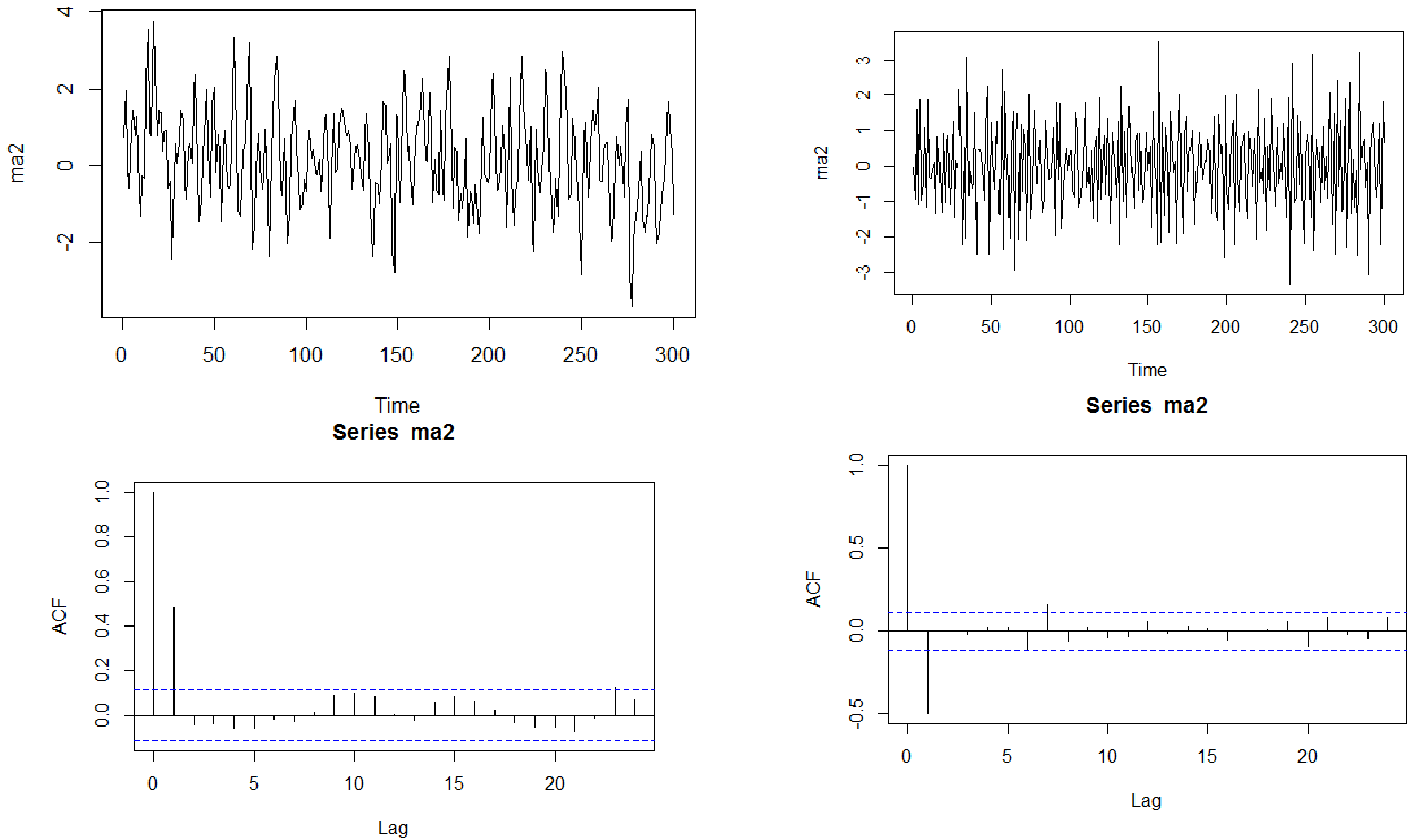


ACF=0, k > 1. Negative correlation at lag 1. Series is choppy.

R Commands:

```
> ma1 <- arima.sim(model=list(ma=c(-0.8)), n=100)
> plot(ma1)
> acf(ma1)
> plot(ARMAacf(ma=c(-0.8), lag.max=40), col="red", type="h", xlab="lag", ylim=c(-.8,1));
abline(h=0)
```

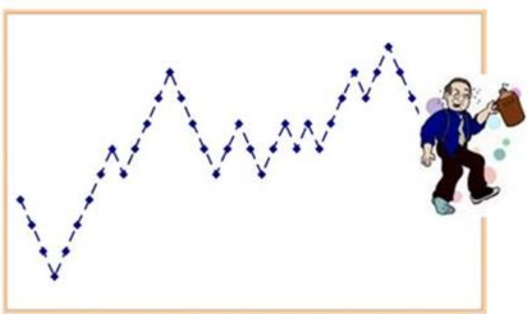
Simulated values/ACF of $X_t = Z_t + 0.95\, Z_{t-1}$ and $X_t = Z_t - 0.95\, Z_{t-1}$



Question: which column corresponds to model $X_t = Z_t - 0.95\, Z_{t-1}$?

Main Concepts of Lecture 2:

- **White Noise Example**
- **Characteristics of Gaussian (Normal) distribution;**
- **Gaussian WN**
- **Strictly and (weakly) stationary time series; their ACVF and ACF**
- **Visual characteristics of nonstationary TS**
- **Examples of calculations of ACVF**
- **Random Walk –example of non stationary TS**
- **Moving Average MA(1) model and its ACF.**



Some simple R commands used to create previous slides

- **To simulate 100 values of MA(1): $X_t = Z_t - 0.95Z_{t-1}$**

```
> ma1 <- arima.sim(model=list(ma=c(-0.95)), n=100)
> plot(ma1)
```

- **To plot ACF for simulated values stored in file ma1:**

```
acf (ma1, type="correlation", plot=T)
```

- **To plot theoretical ACF for this MA(1)**

```
plot(ARMAacf(ma=c(-0.95), lag.max=40), col="red", type="h", xlab="lag",
ylim=c(-.8,1)); abline(h=0)
```

- **Plots with airpass data were produced with the following commands:**

```
> AP=read.table("AIRPASS.txt")
> plot.ts(AP)
> AP <-scan("AIRPASS.txt")                 %read 144 data points
> nt=length(AP)
> fit <- lm(AP ~ as.numeric(1:nt))
> abline(fit)                              %added trend to data plot
> mean(AP)[1] 280.2986
> abline(h=mean(AP), col="red")            %added mean (constant) to data plot)
> acf(AP)                                  % plot acf
```