# Lecture 1

## An Overview of Time Series Analysis

Readings: CC08 Chapter 1; SS17 Chapter 1; BD16 Chapter 1.1 - 1.3

*MATH 8090 Time Series Analysis*
Week 1

Whitney Huang
Clemson University

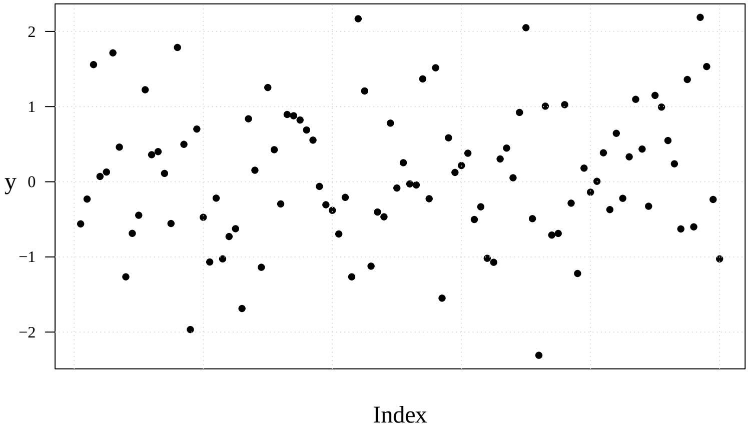
# Agenda

1. **Time Series Data**

2. **Time Series Models**

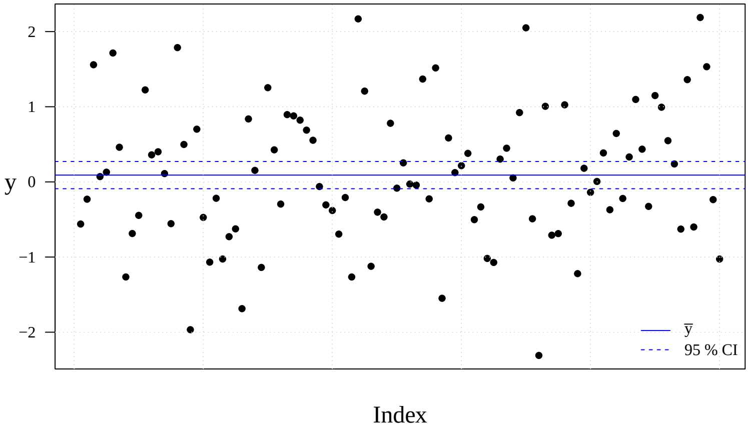
3. **Objectives of Time Series Analysis**

# Statistical Inference from a Sample



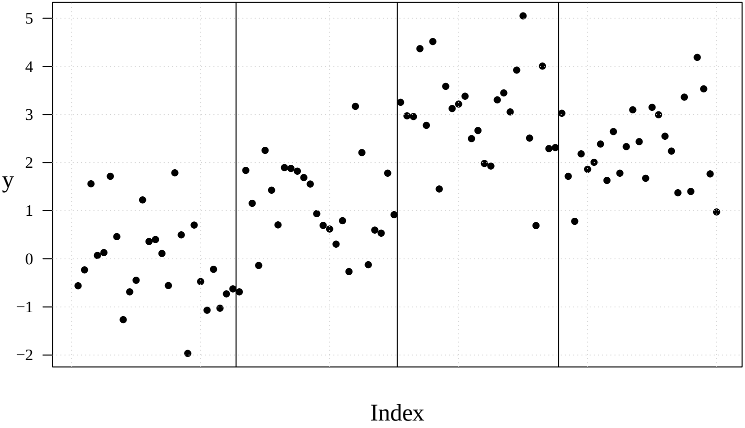

**How can we use a random sample to infer the population mean?**

# Statistical Inference from a Sample



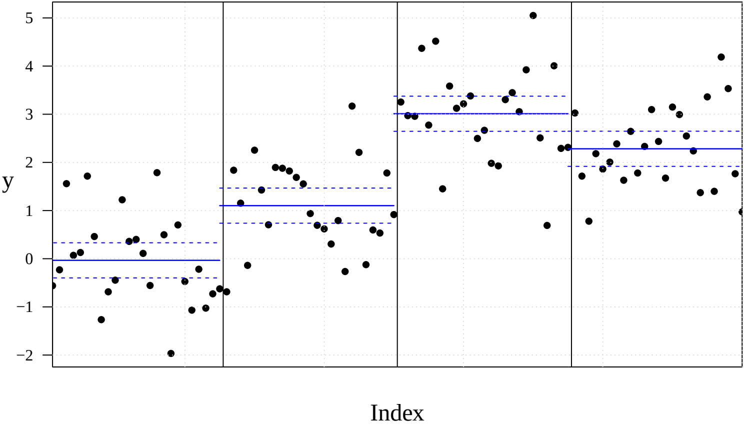

**How do we obtain the point and interval estimates?**

# Statistical Inference for Group Means



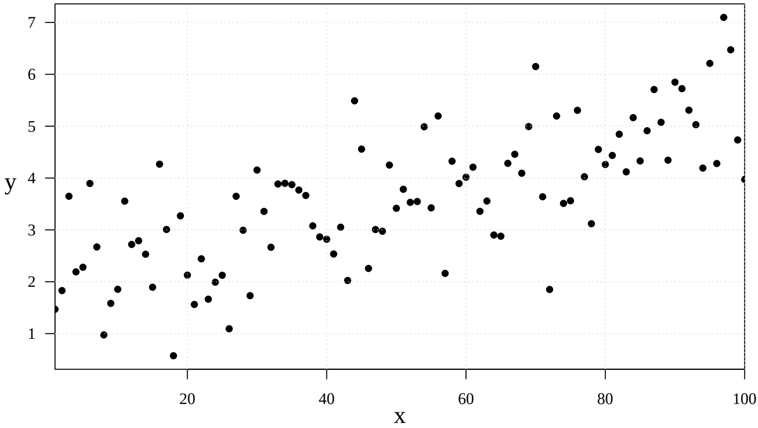

**Are the group means significantly different?**

# Statistical Inference for Group Means



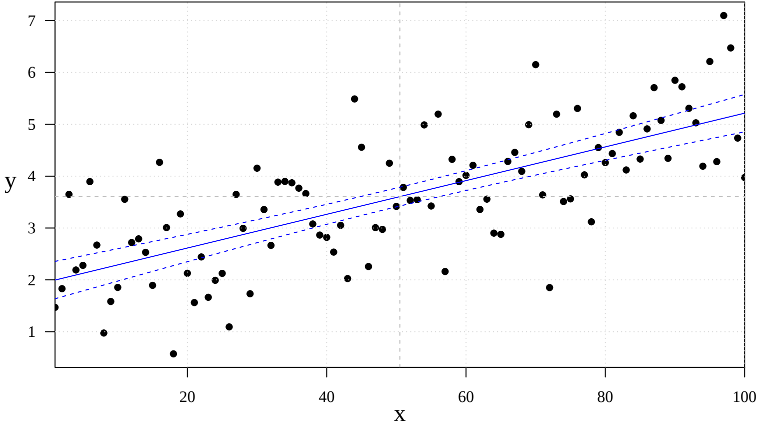

**Which statistical technique is used here?**

# Statistical Inference for Conditional Mean Curve




**How does $y$ change conditionally on $x$?**

## Statistical Inference for Conditional Mean Curve

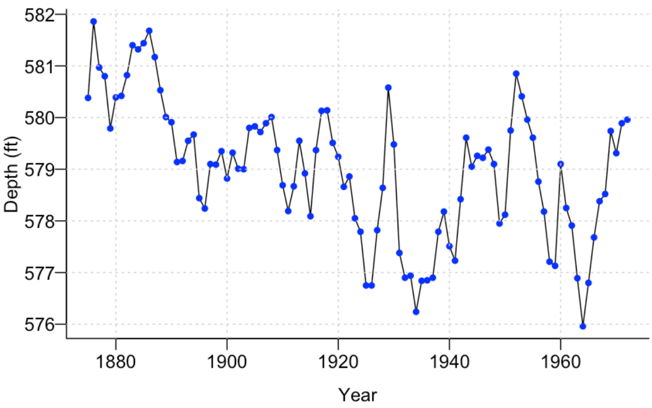



**Which statistical technique is used here?**

# Level of Lake Huron 1875–1972

Annual measurements of the level of Lake Huron in feet.
[Source: Brockwell & Davis, 1991]

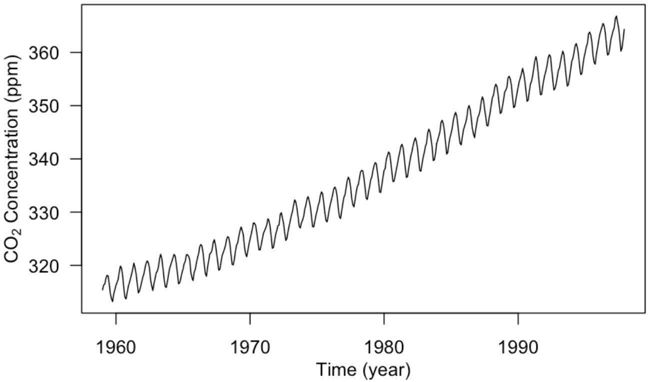
```{r}
par(mar = c(3.2, 3.2, 0.5, 0.5), mgp = c(2, 0.5, 0), bty = "L")
data(LakeHuron)
plot(LakeHuron, ylab = "Depth (ft)", xlab = "Year", las = 1)
points(LakeHuron, cex = 0.8, col = "blue", pch = 16)
grid()
```



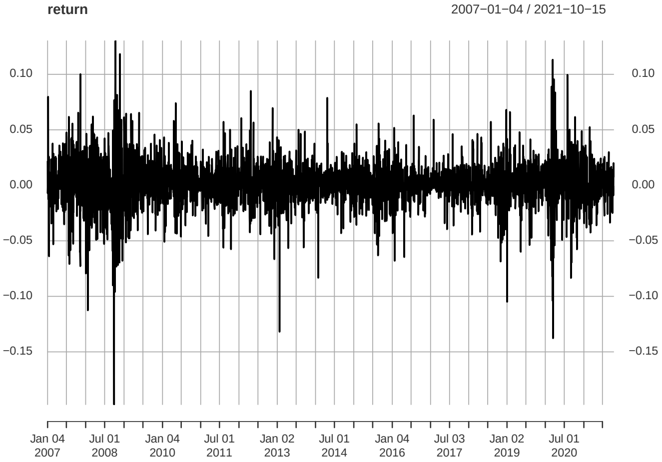
# Mauna Loa Atmospheric $CO_2$ Concentration

Monthly atmospheric concentrations of $CO_2$ at the Mauna Loa Observatory [Source: Keeling & Whorf, Scripps Institution of Oceanography]

```{r}
data(co2)
par(mar = c(3.8, 4, 0.8, 0.6))
plot(co2, las = 1, xlab = "", ylab = "")
mtext("Time (year)", side = 1, line = 2)
mtext(expression(paste("CO"[2], " Concentration (ppm)")), side = 2, line = 2.5)
```



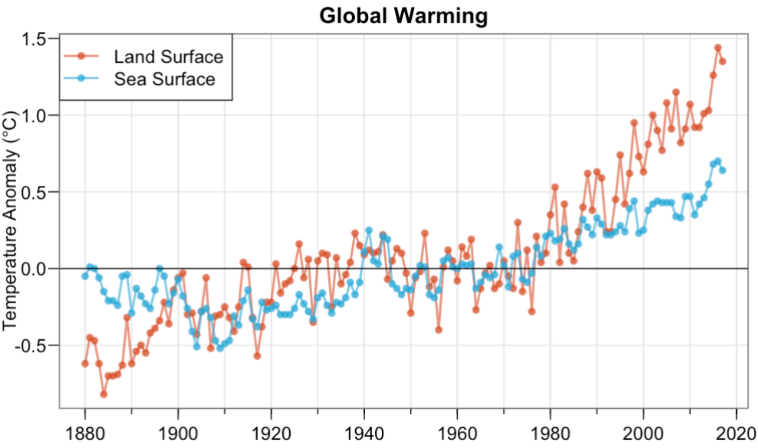
# Apple Stock Log Returns



# Global Annual Temperature Anomalies

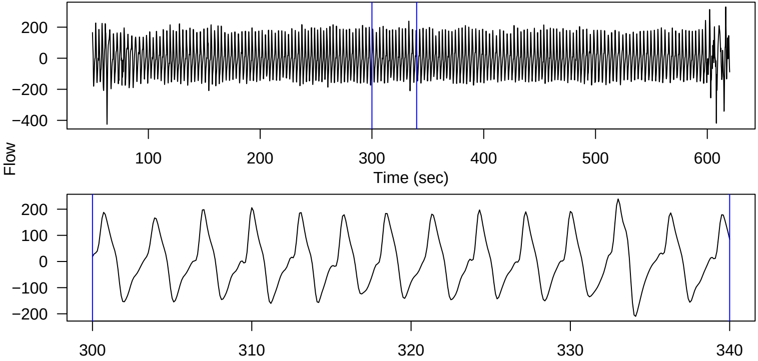


[Source: NASA GISS Surface Temperature Analysis]



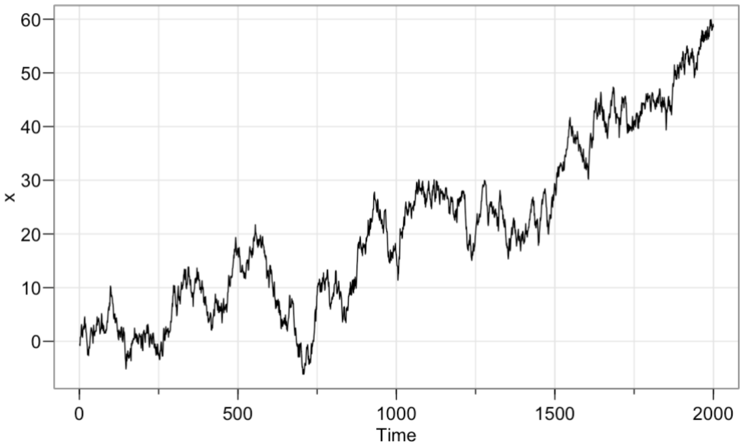
# Sleep Airflow Signal

A “normal” patient’s 10 Hz sleep airflow signal [Source: H. et al. 2022]



# A Simulated Time Series

```{r}
set.seed(123)
w <- rnorm(2000); x <- cumsum(w); tsplot(x, las = 1)
```



# Time Series Data

- A time series is a collection of observations $\{y_t, t \in T\}$ taken sequentially in time $(t)$ with the index set $T$
  - $T = \{0, 1, 2, \cdots, T\} \subset \mathbb{Z} \Rightarrow$ discrete-time time series
  - $T = [0, T] \subset \mathbb{R} \Rightarrow$ continuous-time time series
- A discrete-time time series might be intrinsically discrete or might arise from a underlying continuous-time time series via
  - sampling (e.g., instantaneous wind speed)
  - aggregation (e.g., daily accumulated precipitation amount)
  - extrema (e.g., daily maximum temperature)
- We will focus on dealing with discrete-time real-valued $(Y_t \in \mathbb{R})$ time series in this course

# Exploratory Time Series Analysis

- Start with a time series plot, i.e., to plot $y_t$ versus $t$
- Look at the following:
  - Are there abrupt changes?
  - Are there “outliers”?
  - Is there a need to transform the data?
- Examine the trend, seasonal components, and the “noise” term

## Features of Times Series

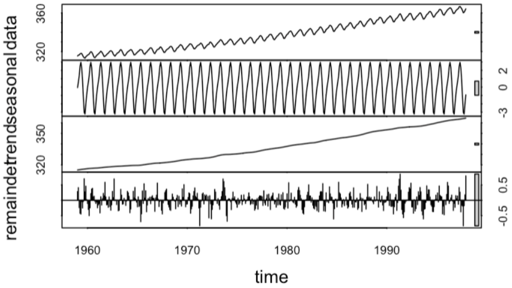
- Trends
  - One can think of trend, $\mu_t$, as continuous changes, usually in the mean, over longer time scales $\Rightarrow$ *"the essential idea of trend is that it shall be smooth"* - [Kendall, 1973]
  - Usually the form of the trend is unknown and needs to be estimated. When the trend is removed, we obtain a detrended series
- Seasonal or periodic components
  - A seasonal component $s_t$ constantly repeats itself in time, i.e., $s_t = s_{t+kd}$
  - We need to estimate the form and/or the period $d$ of the seasonal component to deseasonalize the series
- The "noise" process
  - The noise process, $\eta_t$, is the component that is neither trend nor seasonality
  - We will focus on finding plausible (typically stationary) statistical models for this process

## Combining Trend, Seasonality, and Noise Together

There are two commonly used approaches

- Additive model:

$$y_t = \mu_t + s_t + \eta_t, \quad t = 1, \cdots, T$$




- Multiplicative model:

$$y_t = \mu_t s_t \eta_t, \quad t = 1, \cdots, T$$

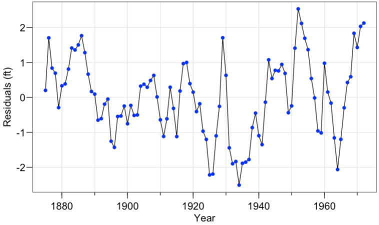
If all $\{y_t\}$ are positive then we obtain the additive model by taking logarithms:

$$\log y_t = \log \mu_t + \log s_t + \log \eta_t, \quad t = 1, \cdots, T$$

# Time Series Models

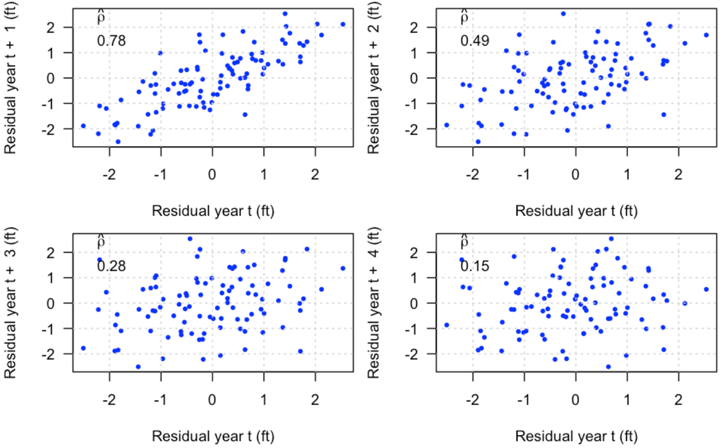
# Lake Huron Time Series

- Time series analysis is the area of statistics which deals with the analysis of dependency between different observations (typically $\{\eta_t\}$)
- Some key features of the Lake Huron time series:
  - decreasing trend
  - some “random” fluctuations around the decreasing trend
- We extract the “noise” component by assuming a linear trend



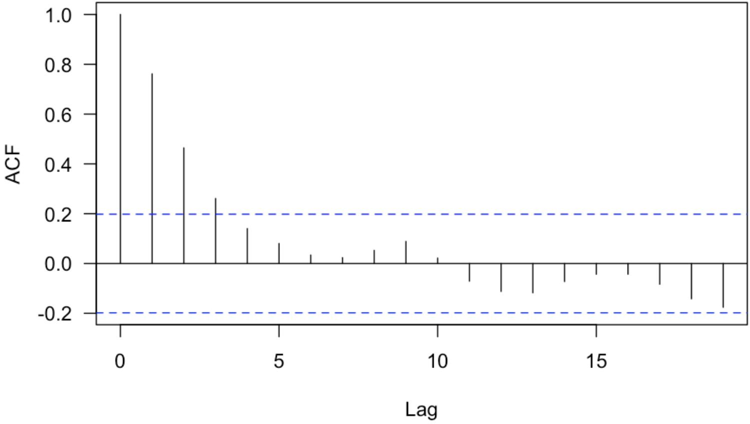
# Exploring the Temporal Dependence Structure of $\{\eta_t\}$

$\{\eta_t\}$ exhibits a temporal dependence structure, meaning that nearby (in time) values tend to be more alike than those far apart. To see this, let's create a few time-lag plots



# Further Exploration of the Temporal Dependence Structure

Let's plot the correlation as a function of the time lag




In a few weeks we will learn how to use this information to suggest an appropriate model

## Time Series Models

- A time series model is a probabilistic model for how the data $\{y_t\}$ may have been generated

## Time Series Models

- A time series model is a probabilistic model for how the data $\{y_t\}$ may have been generated
- More formally, it is a probability model for $\{Y_t : t \in T\}$, a collection of random variables indexed in time

## Time Series Models

- A time series model is a probabilistic model for how the data $\{y_t\}$ may have been generated
- More formally, it is a probability model for $\{Y_t : t \in T\}$, a collection of random variables indexed in time
- We keep models simple by assuming stationarity $\Rightarrow$ distributional properties depend only on time lags, not on specific time points

# Time Series Models

- A time series model is a probabilistic model for how the data $\{y_t\}$ may have been generated
- More formally, it is a probability model for $\{Y_t : t \in T\}$, a collection of random variables indexed in time
- We keep models simple by assuming stationarity $\Rightarrow$ distributional properties depend only on time lags, not on specific time points
- Most time series are not stationary; we remove or model non-stationarity (e.g., de-trending, de-seasonalizing) to obtain a stationary component $\{\eta_t\}$. Typically, we assume second-order stationarity:

$$\mathrm{E}[\eta_t] = 0 \text{ for all } t, \text{ and}$$
$$\mathrm{Cov}(\eta_t, \eta_{t'}) = \gamma(t' - t) = \mathrm{Cov}(\eta_{t+s}, \eta_{t'+s})$$

# Objectives of Time Series Analysis

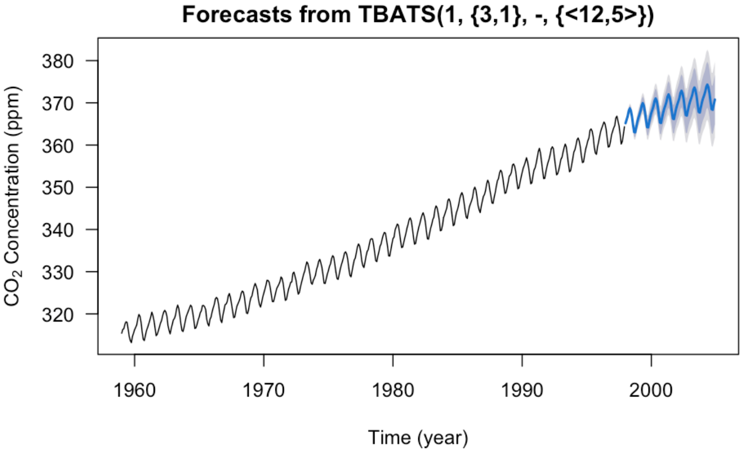
# Some Objectives of Time Series Analysis



- **Modeling**: find a statistical model that adequately explains the observed time series
- For example, identify a model that accounts for the fact that Lake Huron depths are correlated across different years and exhibit a decreasing long-term trend
- The fitted model can be used for further statistical inference, for instance, to answer questions such as: Is there evidence of a decreasing trend in Lake Huron depths?

## Some Objectives of Time Series Analysis, Cont'd

Forecasting is perhaps the most common objective. One observes a time series of given length and wishes to **predict** or **forecast** future values based on the observed data



# Some Objectives of Time Series Analysis, Cont'd

- **Adjustment**: e.g., seasonal adjustment, where the seasonal component is estimated and removed to better reveal the underlying trend
- **Simulation**: use a time series model (that adequately describes a physical process) as a surrogate to *simulate repeatedly and approximate the process's behavior*
- **Control**: adjust input (control) parameters to make the series conform more closely to a given standard (many examples come from statistical quality control)