Statistics and Data Analysis
Unit 06 – Lecture 01 Notes
Time-series Concepts (Trend, Seasonality, Autocorrelation)

Tofik Ali

February 17, 2026

Topic

Time series basics: components and autocorrelation.

How to Use These Notes

These notes are written for students who are seeing the topic for the first time. They follow the slide order, but add the missing 'why', interpretation, and common mistakes. If you get stuck, look at the worked exercises and then run the Python demo.

Course repository (slides, demos, datasets): `https://github.com/tali7c/Statistics-and-Data-Analysis`

Time Plan (55 minutes)

- 0–10 min: Attendance + recap of previous lecture
- 10–35 min: Core concepts (this lecture's sections)
- 35–45 min: Exercises (solve 1–2 in class, rest as practice)
- 45–50 min: Mini demo + interpretation of output
- 50–55 min: Buffer / wrap-up (leave 5 minutes early)

Slide-by-slide Notes

Title Slide

State the lecture title clearly and connect it to what students already know. Tell students what they will be able to do by the end (not just what you will cover).

Quick Links / Agenda

Explain the structure of the lecture and where the exercises and demo appear.

- Overview

- Components
- Autocorrelation
- Exercises
- Demo
- Summary

Learning Outcomes

- Define time series and why order matters
- Identify trend, seasonality, and noise
- Explain autocorrelation (intuition)
- Explain why random shuffling breaks time series analysis

Why these outcomes matter. **Correlation** measures the strength of a linear association between two variables. It is symmetric (no X/Y direction) and does not imply causation. Outliers can inflate or hide correlation, so always look at the scatter plot. A **time series** is data indexed by time (daily sales, hourly sensor readings). The key difference from 'normal' data is that order matters and observations are often correlated over time. Many standard ML assumptions (IID, random split) break for time series.

Components: Key Points

- Trend: long-term movement
- Seasonality: repeating pattern
- Noise: irregular fluctuations

Explanation. **Trend** is a long-term upward or downward movement. Trend changes the mean over time, which often creates non-stationarity. Many forecasting models handle trend by differencing or by explicitly modeling trend. **Seasonality** is a repeating pattern with a fixed period (weekly, monthly, yearly). You must account for it; otherwise forecasts systematically miss repeating rises/falls. Seasonal differencing and SARIMA are common tools.

Autocorrelation: Key Points

- Correlation with past values
- Important for AR/MA/ARIMA models
- Shows persistence of shocks

Explanation. **Correlation** measures the strength of a linear association between two variables. It is symmetric (no X/Y direction) and does not imply causation. Outliers can inflate or hide correlation, so always look at the scatter plot. **Autocorrelation** measures how strongly the series relates to its past values. Positive autocorrelation means high values tend to follow high values; negative means alternation. Autocorrelation is the reason AR/MA/ARIMA models work at all.

Exercises (with Solutions)

Attempt the exercise first, then compare with the solution. Focus on interpretation, not only arithmetic.

Exercise 1: Order matters

Why should train/test split be chronological for time series?

Solution

- To avoid future-to-past leakage.

Walkthrough. **Data leakage** happens when information from the future or from the test set influences training. Typical examples: scaling before splitting, using target-related features, or using random splits for time series. Leakage can produce very good-looking accuracy that disappears in real deployment. A **time series** is data indexed by time (daily sales, hourly sensor readings). The key difference from 'normal' data is that order matters and observations are often correlated over time. Many standard ML assumptions (IID, random split) break for time series.

Exercise 2: Seasonality example

Give one seasonal pattern in campus data.

Solution

- Weekly cafe sales (weekday vs weekend), etc.

Walkthrough. **Seasonality** is a repeating pattern with a fixed period (weekly, monthly, yearly). You must account for it; otherwise forecasts systematically miss repeating rises/falls. Seasonal differencing and SARIMA are common tools.

Exercise 3: Autocorr meaning

If lag-1 autocorrelation is strong positive, what does it suggest?

Solution

- Values tend to persist from one step to next.

Walkthrough. **Correlation** measures the strength of a linear association between two variables. It is symmetric (no X/Y direction) and does not imply causation. Outliers can inflate or hide correlation, so always look at the scatter plot. **Autocorrelation** measures how strongly the series relates to its past values. Positive autocorrelation means high values tend to follow high values; negative means alternation. Autocorrelation is the reason AR/MA/ARIMA models work at all.

Mini Demo (Python)

Run from the lecture folder:

```
python demo/demo.py
```

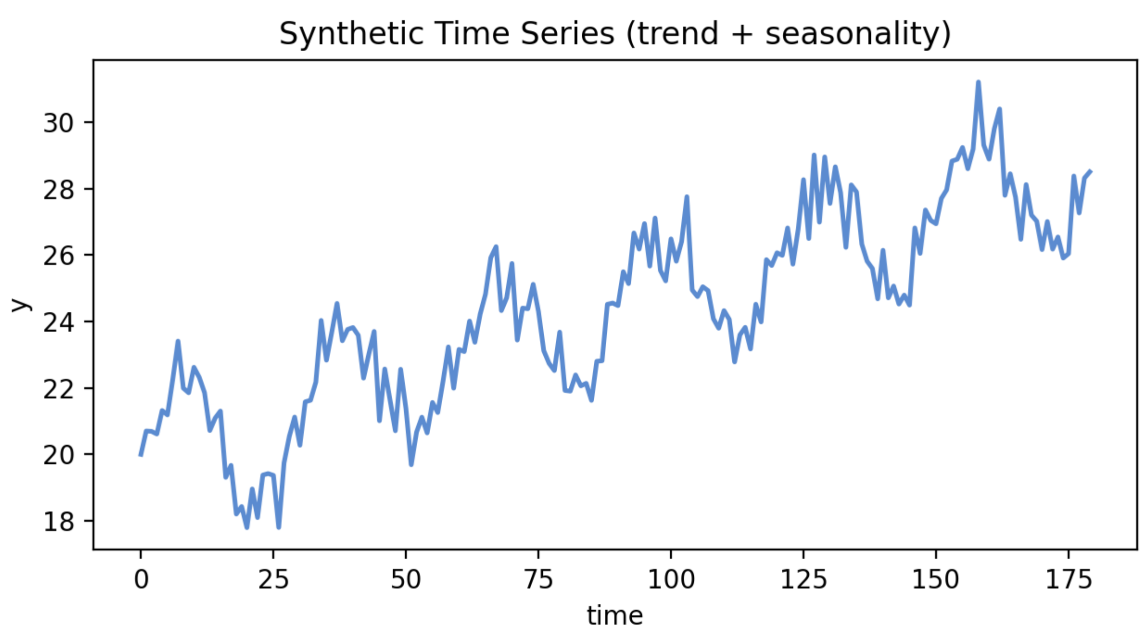
Output files:

- `images/demo.png`
- `data/results.txt`

What to show and say.

- Generates a time series with trend/seasonality and plots it.
- Shows how autocorrelation makes random shuffling invalid.
- Use it to motivate chronological splits and forecasting mindset.

Demo Output (Example)



Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

Exit Question

In one sentence: what is seasonality and why does it matter for forecasting?

Suggested answer (for revision). Seasonality is a repeating pattern with a fixed period; it matters because forecasts must reproduce these cycles or they will be systematically wrong.

References

- Montgomery, D. C., & Runger, G. C. *Applied Statistics and Probability for Engineers*, Wiley.
- Devore, J. L. *Probability and Statistics for Engineering and the Sciences*, Cengage.
- McKinney, W. *Python for Data Analysis*, O'Reilly.

Appendix: Slide Deck Content (Reference)

The material below is a reference copy of the slide deck content. Exercise solutions are explained in the main notes where applicable.

Title Slide

https://github.com/tali7c/Statistics-and-Data-Analysis

Quick Links

Overview **Components** **Autocorrelation** **Exercises** **Demo** **Summary**

Agenda

- Overview
- Components
- Autocorrelation
- Exercises
- Demo
- Summary

Learning Outcomes

- Define time series and why order matters
- Identify trend, seasonality, and noise
- Explain autocorrelation (intuition)
- Explain why random shuffling breaks time series analysis

Components: Key Points

- Trend: long-term movement
- Seasonality: repeating pattern
- Noise: irregular fluctuations

Autocorrelation: Key Points

- Correlation with past values
- Important for AR/MA/ARIMA models
- Shows persistence of shocks

Exercise 1: Order matters

Why should train/test split be chronological for time series?

Solution 1

- To avoid future-to-past leakage.

Exercise 2: Seasonality example

Give one seasonal pattern in campus data.

Solution 2

- Weekly cafe sales (weekday vs weekend), etc.

Exercise 3: Autocorr meaning

If lag-1 autocorrelation is strong positive, what does it suggest?

Solution 3

- Values tend to persist from one step to next.

Mini Demo (Python)

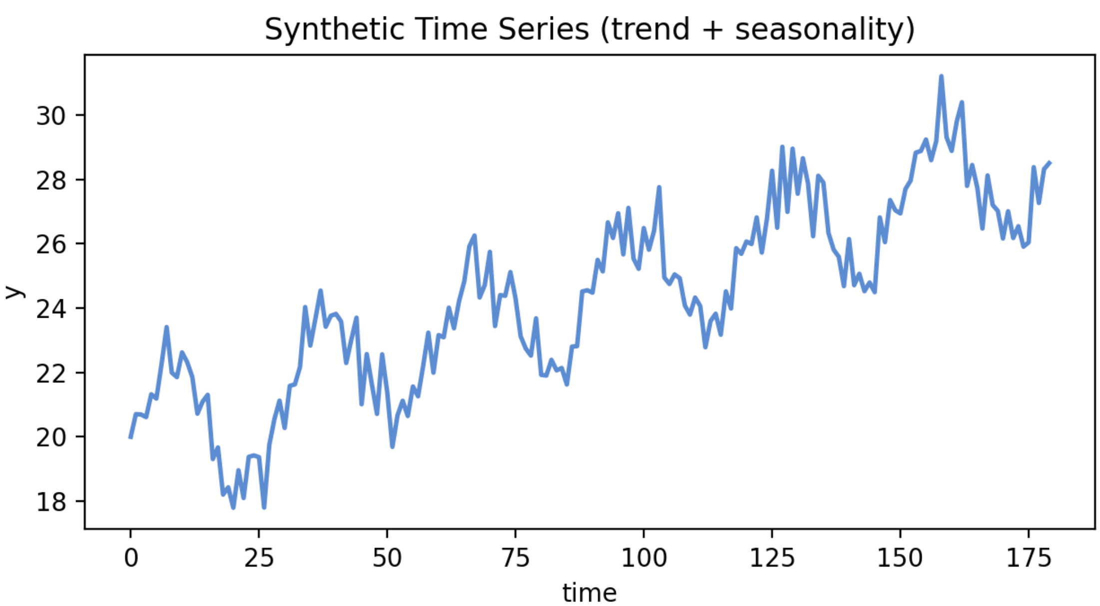
Run from the lecture folder:

```
python demo/demo.py
```

Outputs:

- `images/demo.png`
- `data/results.txt`

Demo Output (Example)



Summary

- Key definitions and the main formula.
- How to interpret results in context.
- How the demo connects to the theory.

Exit Question

In one sentence: what is seasonality and why does it matter for forecasting?