

# Statistics and Data Analysis

## Unit 06 – Lecture 08 Notes

### Diagnostics and SARIMA Models

Tofik Ali

February 17, 2026

#### Topic

Time-series diagnostics and SARIMA seasonal modeling (overview).

#### 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

- Diagnostics
- SARIMA
- Exercises
- Demo
- Summary

## Learning Outcomes

- Explain why diagnostics are needed after fitting
- Recognize residual autocorrelation as model issue
- Explain SARIMA seasonal terms at a high level
- Choose seasonal period  $s$  (weekly/monthly/yearly)

**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 **residual** is  $y - \hat{y}$ . Residual plots tell you what the model failed to explain. Patterns in residuals (trend, curvature, changing variance) are warnings that your model form is inadequate or assumptions are violated.

## Diagnostics: Key Points

- Residuals should look like white noise
- Check residual ACF
- Check stability of variance

**Explanation.** A **residual** is  $y - \hat{y}$ . Residual plots tell you what the model failed to explain. Patterns in residuals (trend, curvature, changing variance) are warnings that your model form is inadequate or assumptions are violated. **ACF** shows correlation of the series with its lagged versions. It helps identify MA- like behavior and seasonality. **PACF** shows the correlation at a lag after removing shorter-lag effects and helps identify AR-like behavior.

## SARIMA: Key Points

- ARIMA( $p,d,q$ )  $\times$  (P,D,Q, $s$ )
- Seasonal differencing  $D$
- $s$  is the seasonal period

**Explanation.** **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. **Differencing** transforms a series by subtracting the previous value:  $y_t - y_{t-1}$ . It removes trend and can help achieve stationarity. Over-differencing can add noise, so use the smallest differencing order that works. **SARIMA** extends ARIMA by adding seasonal AR/MA terms and seasonal differencing. It is used when patterns repeat every  $s$  steps (e.g., weekly seasonality with daily data). Diagnostics on residuals are important: residuals should look like white noise.

## Exercises (with Solutions)

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

### Exercise 1: Residual goal

After fitting, what should residuals look like ideally?

#### Solution

- White noise: no pattern, no autocorrelation.

**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. A **residual** is  $y - \hat{y}$ . Residual plots tell you what the model failed to explain. Patterns in residuals (trend, curvature, changing variance) are warnings that your model form is inadequate or assumptions are violated.

### Exercise 2: Seasonal period

Daily data with weekly seasonality: what is  $s$ ?

#### Solution

- $s=7$

**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: Why seasonal

Why add seasonal terms?

#### Solution

- To capture repeating seasonal dependence.

**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.

## Mini Demo (Python)

Run from the lecture folder:

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

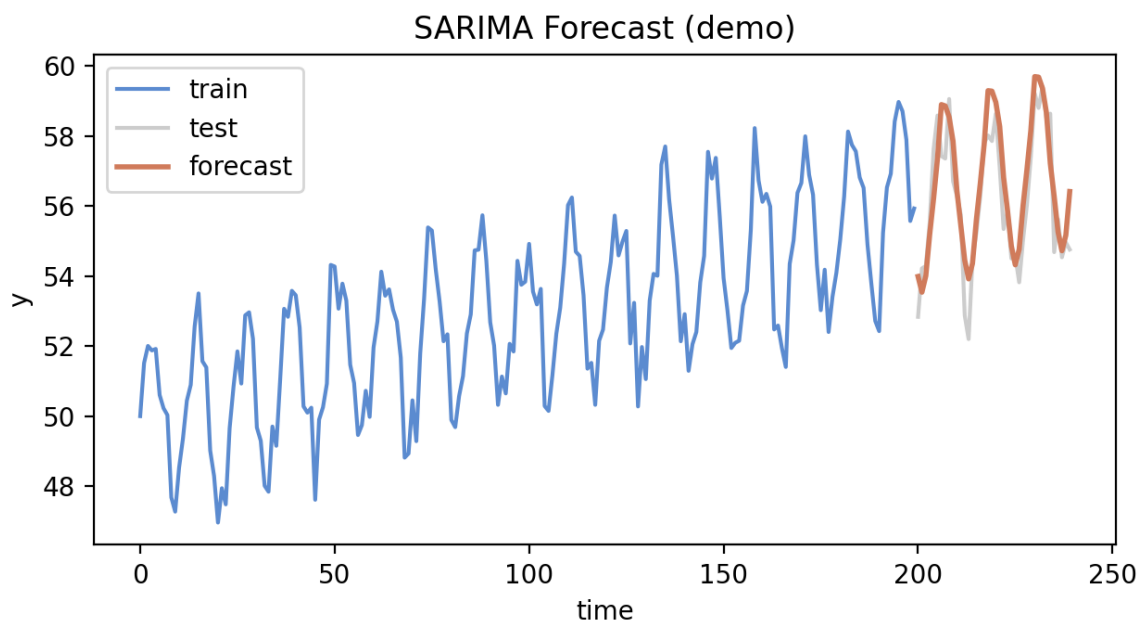
Output files:

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

## What to show and say.

- Fits a seasonal model idea and checks residual diagnostics conceptually.
- Shows how seasonal period  $s$  (e.g., 7 for weekly) enters the model.
- Use residual ACF idea to explain when the model is still missing structure.

## 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

What is one residual symptom that suggests your model is inadequate?

**Suggested answer (for revision).** Residual autocorrelation (or visible patterns) suggests the model is missing structure; add terms/seasonality until residuals resemble white noise.

## 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**

## Quick Links

[Overview](#) [Diagnostics](#) [SARIMA](#) [Exercises](#) [Demo](#) [Summary](#)

## Agenda

- Overview
- Diagnostics
- SARIMA
- Exercises
- Demo
- Summary

## Learning Outcomes

- Explain why diagnostics are needed after fitting
- Recognize residual autocorrelation as model issue
- Explain SARIMA seasonal terms at a high level
- Choose seasonal period  $s$  (weekly/monthly/yearly)

## Diagnostics: Key Points

- Residuals should look like white noise
- Check residual ACF
- Check stability of variance

## SARIMA: Key Points

- $ARIMA(p,d,q) \times (P,D,Q,s)$
- Seasonal differencing  $D$
- $s$  is the seasonal period

## Exercise 1: Residual goal

After fitting, what should residuals look like ideally?

## Solution 1

- White noise: no pattern, no autocorrelation.

## Exercise 2: Seasonal period

Daily data with weekly seasonality: what is  $s$ ?

### Solution 2

- $s=7$

## Exercise 3: Why seasonal

Why add seasonal terms?

### Solution 3

- To capture repeating seasonal dependence.

## Mini Demo (Python)

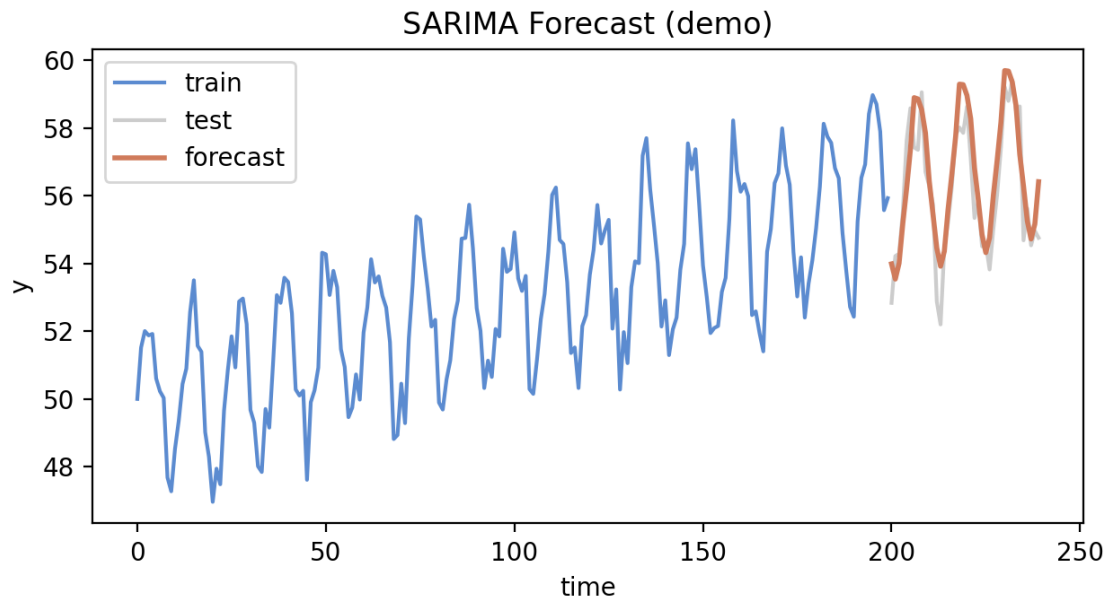
Run from the lecture folder:

```
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**

What is one residual symptom that suggests your model is inadequate?