

Statistics and Data Analysis

Unit 06 – Lecture 04 Notes

Tofik Ali

February 14, 2026

Topic

ARIMA models and forecasting workflow (overview).

Learning Outcomes

- Define ARIMA(p,d,q) at a high level
- Explain differencing (d) to remove trend
- Explain p and q meaning (AR and MA orders)
- Describe time-based train/test split for forecasting

Detailed Notes

These notes are designed to be read alongside the slides. They expand each slide bullet into plain-language explanations, small worked examples, and common pitfalls. When a formula appears, emphasize (1) what each symbol means, (2) the assumptions needed to use it, and (3) how to interpret the final number in the problem context.

ARIMA

- p: AR order
- d: differencing order
- q: MA order

Differencing

- First difference: $y_t - y_{t-1}$
- Often stabilizes mean
- Over-differencing adds noise

Exercises (with Solutions)

Exercise 1: Meaning of d

What does $d=1$ mean?

Solution

- First differencing once.

Exercise 2: Chronological split

Why not random split in time series?

Solution

- Random split leaks future information.

Exercise 3: Trend fix

Series has strong upward trend. Name one simple step.

Solution

- First differencing.

Exit Question

Why do we check residuals after fitting an ARIMA model?

Demo (Python)

Run from the lecture folder:

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

Output files:

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

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.