

Statistics and Data Analysis

Unit 03 – Lecture 01 Notes

Tofik Ali

February 14, 2026

Topic

Population vs sample; sampling techniques; estimation; confidence intervals (overview).

Learning Outcomes

- Differentiate population parameters and sample statistics
- Explain sampling bias vs random error
- Describe common sampling methods (SRS, stratified, cluster)
- Compute and interpret a basic confidence interval for a mean

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.

Sampling

- Population vs sample
- Bias vs random error
- Representative sampling matters

Confidence Intervals

- Interpretation: long-run coverage
- Width depends on n and variability
- CIs support decision-making with uncertainty

Exercises (with Solutions)

Exercise 1: Parameter vs Statistic

Give one example of a parameter and one of a statistic.

Solution

- Parameter: population mean
- Statistic: sample mean

Exercise 2: CI Interpretation

In one sentence, what does a 95% CI mean (correctly)?

Solution

- About 95% of such intervals contain the true mean in repeated sampling.

Exercise 3: Bias Scenario

Why does convenience sampling create bias?

Solution

- Because some groups are over/under-represented systematically.

Exit Question

If your CI is too wide, what two actions reduce its width (without cheating)?

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.