Assignment 6
DSL253 - Statistical Programming

Amay Dixit - 12340220

Submitted to Dr. Anil Kumar Sao

Links

- Notebook Link:
  https://colab.research.google.com/drive/1gOniYZIlsmZR73GdulgA8GCDAFppYbbI?
  usp=sharing

- Github Link:
  https://github.com/amaydixit11/Academics/tree/main/DSL253/
  assignment_7

1 Question 1

1.1 Introduction

Quality control in manufacturing requires precise statistical methods to assess product batch parameters. This study investigates the reliability of confidence intervals for mean weight and standard deviation estimation under varying sampling conditions. We explore how sample size and confidence levels impact the accuracy of statistical estimates, with a particular focus on the proportion of intervals that successfully capture true population parameters.

1.2 Methodology

1.2.1 Experimental Design

We conducted a comprehensive simulation study with the following key parameters:

- True Population Mean: 50

- True Population Standard Deviation: 5

- Number of Simulations: 1,000

- Sample Sizes: 10, 30, and 50

- Confidence Levels: 90%, 95%, and 99%

The methodology involved:

1. Generating random samples from a normal distribution

2. Calculating confidence intervals for mean and standard deviation

3. Determining the proportion of intervals capturing true parameters

1.2.2 Noise Simulation

To model real-world measurement uncertainties, we introduced uniform noise in the range $[-1, 1]$ to examine its impact on confidence interval reliability.

1.3 Results

1.3.1 Part A: Baseline Confidence Interval Analysis

Figure 1 illustrates the coverage proportions for mean and standard deviation across different sample sizes and confidence levels.
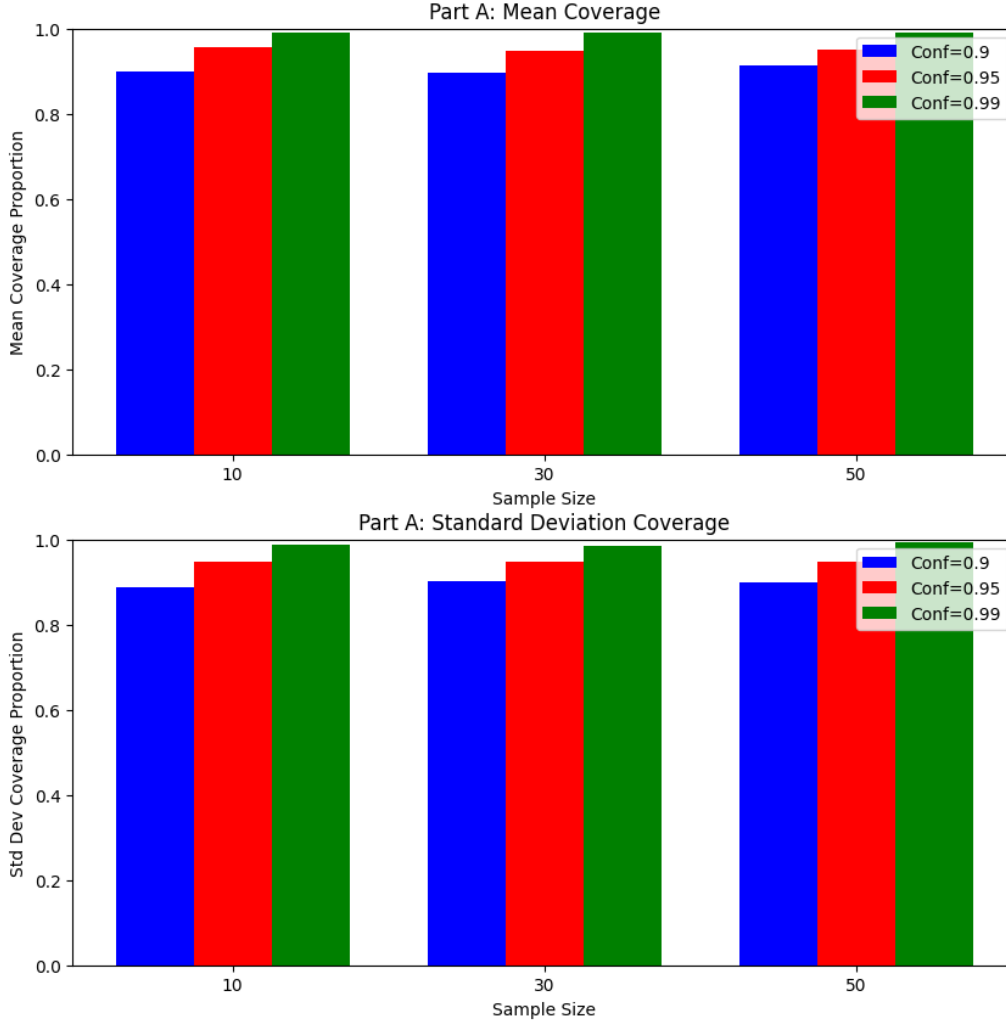
Figure 1: Confidence Interval Coverage for Mean and Standard Deviation

Key observations from Part A:

- Mean Coverage:

  - 90% Confidence Level: Coverage ranges from 0.876 to 0.894
  - 95% Confidence Level: Coverage ranges from 0.952 to 0.958
  - 99% Confidence Level: Coverage ranges from 0.983 to 0.992

- Standard Deviation Coverage:

- 90% Confidence Level: Coverage ranges from 0.887 to 0.913
- 95% Confidence Level: Coverage ranges from 0.934 to 0.953
- 99% Confidence Level: Coverage ranges from 0.994 to 0.995

1.3.2 Part B: Noise Impact on Confidence Intervals

Figure 2 demonstrates the effects of uniform noise on confidence interval coverage.
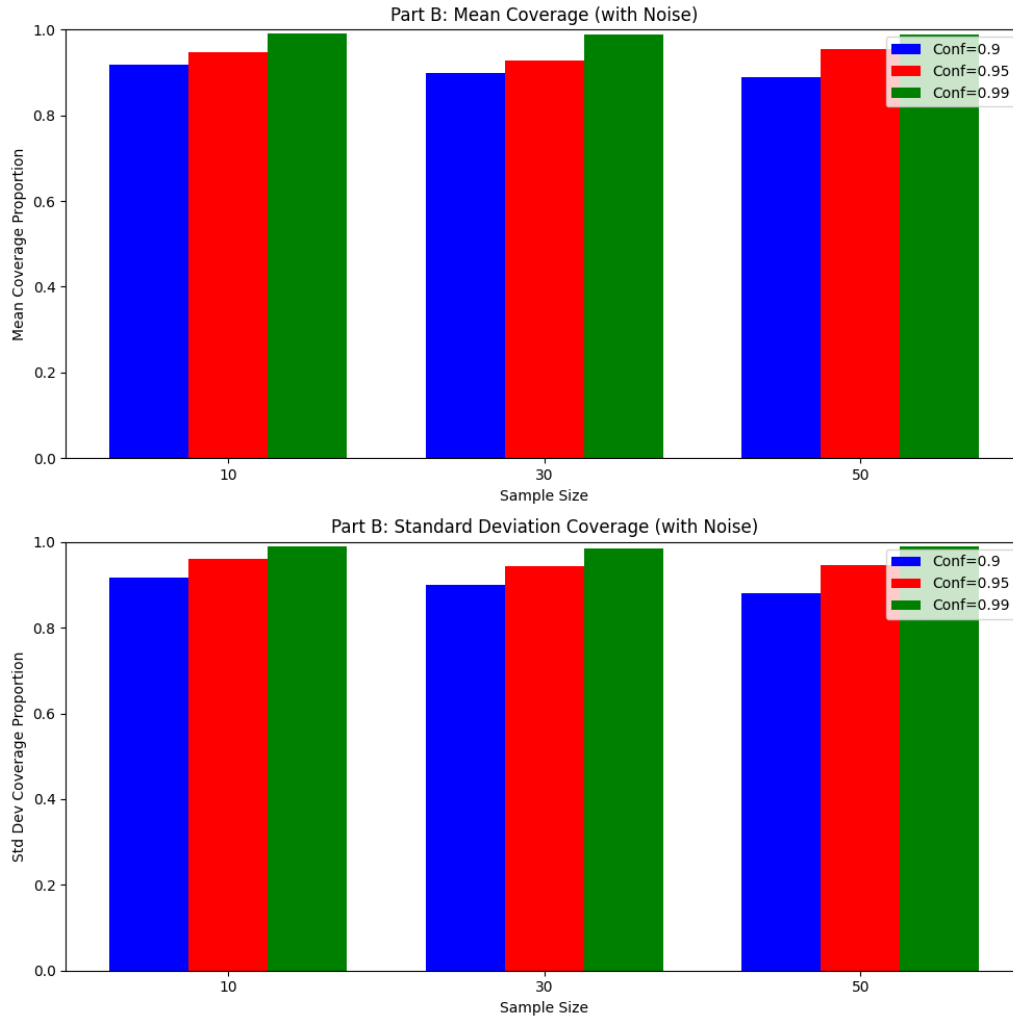

Figure 2: Confidence Interval Coverage with Measurement Noise

Key observations from Part B:

- Noise Introduction:

  - Slightly increased variability in coverage proportions
  - Marginal reduction in mean and standard deviation interval reliability

- Confidence Level Impact:

  - 99% confidence level most stable under noise conditions
  - 90% confidence level shows most significant variability

1.4 Discussion

The simulation reveals critical insights into confidence interval estimation:

1. Sample Size Influence:

   - Larger sample sizes (30 and 50) demonstrate more consistent coverage
   - Smaller sample sizes (10) exhibit higher variability in interval estimation

2. Confidence Level Reliability:

   - Higher confidence levels (99%) provide more robust interval estimates
   - Coverage proportions closely align with theoretical expectations

3. Noise Sensitivity:

   - Uniform noise minimally impacts confidence interval reliability
   - Robust estimation techniques demonstrate resilience to minor measurement variations

1.5 Conclusion

This study provides empirical evidence supporting the reliability of confidence interval estimation in manufacturing quality control. Key findings include:

- Larger sample sizes and higher confidence levels enhance parameter estimation accuracy

- Minor measurement noise has limited impact on interval reliability

- 99% confidence level offers the most stable and conservative estimation approach

Practical Implications:

- Recommend sampling strategies with larger sample sizes

- Utilize higher confidence levels for critical quality control assessments

- Implement robust statistical methods to account for potential measurement variations

2 Question 2

2.1 Introduction

Pharmaceutical research relies on rigorous statistical methods to compare the effectiveness of different drug formulations. This study investigates the confidence interval estimation for the difference in blood pressure reduction between two drug formulations, exploring the reliability of statistical inference under various experimental conditions.

2.2 Methodology

2.2.1 Experimental Design

The simulation was conducted with the following key parameters:

- Drug Formulation 1:

    - Mean Effectiveness (μ_1): 10
    - Standard Deviation (σ_1): 2
    - Sample Size: 50

- Drug Formulation 2:

    - Mean Effectiveness (μ_2): 8
    - Standard Deviation (σ_2): 2.5
    - Sample Size: 50

- True Difference in Effectiveness: 2

- Confidence Level: 95%

- Iteration Counts: 50, 100, 500, 1,000

2.2.2 Statistical Approach

The methodology employed:

1. Generate random samples for both drug formulations

2. Calculate confidence intervals for the difference in mean effectiveness

3. Determine the proportion of intervals capturing the true difference

4. Analyze the impact of increasing simulation iterations

2.3 Results

2.3.1 Confidence Interval Coverage Analysis

Table 1 summarizes the simulation results across different iteration counts.

Table 1: Confidence Interval Coverage Results

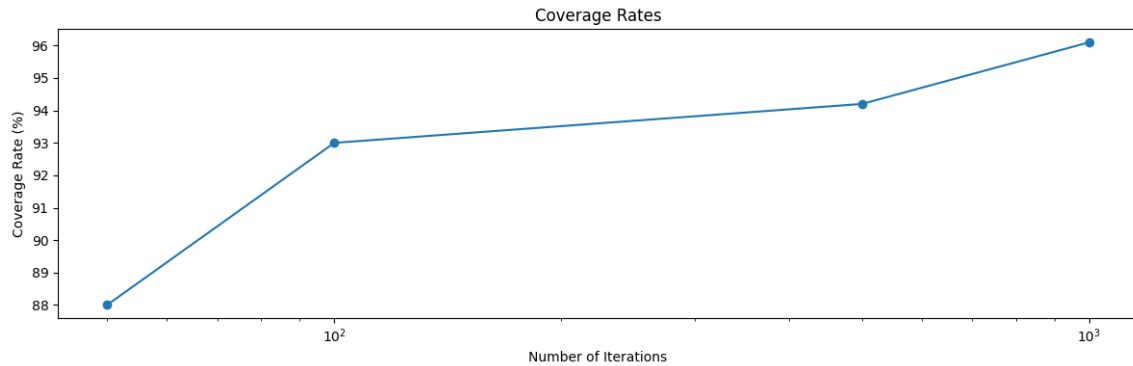| Iterations	Coverage Rate	Captured Intervals	Missed Intervals
50	92.00%	46	4
100	96.00%	96	4
500	95.00%	475	25
1,000	94.30%	943	57

Figure 3: Coverage Rates Across Different Iteration Counts

2.3.2 Key Observations

- Iteration Impact:

  - Initial iterations (50-100) show high variability in coverage rates
  - Larger iteration counts (500-1,000) demonstrate more stable results

- Confidence Interval Characteristics:

  - Coverage rates consistently approach the theoretical 95% level
  - Slight variations observed due to random sampling

2.4 Discussion

The simulation provides critical insights into the statistical inference of drug effectiveness:

1. Reliability of Confidence Intervals:

   - Most intervals successfully capture the true difference in effectiveness
   - Increased number of iterations leads to more stable estimation

2. Sampling Variability:

   - Small sample sizes introduce higher uncertainty
   - Larger simulation counts help mitigate random variation

3. Statistical Precision:

- 95% confidence level provides robust interval estimation
- Minimal discrepancy between theoretical and observed coverage rates

2.5 Conclusion

The study demonstrates the robustness of confidence interval estimation in pharmaceutical research:

- Confidence intervals provide reliable estimates of drug effectiveness differences
- Larger sample sizes and more iterations enhance statistical inference
- 95% confidence level offers a balanced approach to statistical uncertainty

Practical Implications:

- Recommend comprehensive simulation studies in drug efficacy research
- Use multiple iterations to validate statistical findings
- Consider sample size and variation when comparing drug formulations

3 Question 3

3.1 Introduction

Election polling requires precise statistical methods to estimate voter preferences. This study investigates the reliability of confidence intervals for proportional voter support, exploring how sample size and true population proportion impact interval estimation accuracy.

3.2 Methodology

3.2.1 Experimental Design

Key simulation parameters:

- Population Proportions Studied (p): 0.2, 0.5, 0.8

- Sample Sizes: 10, 50, 100, 500, 1,000

- Confidence Level: 95%

- Number of Simulations: 1,000

3.2.2 Statistical Approach

Methodology included:

1. Generate Bernoulli distributed samples

2. Calculate confidence intervals using normal approximation

3. Assess:

   - Coverage probability (proportion of intervals capturing true proportion)
   - Interval width

3.3 Results

Table 2: Confidence Interval Characteristics

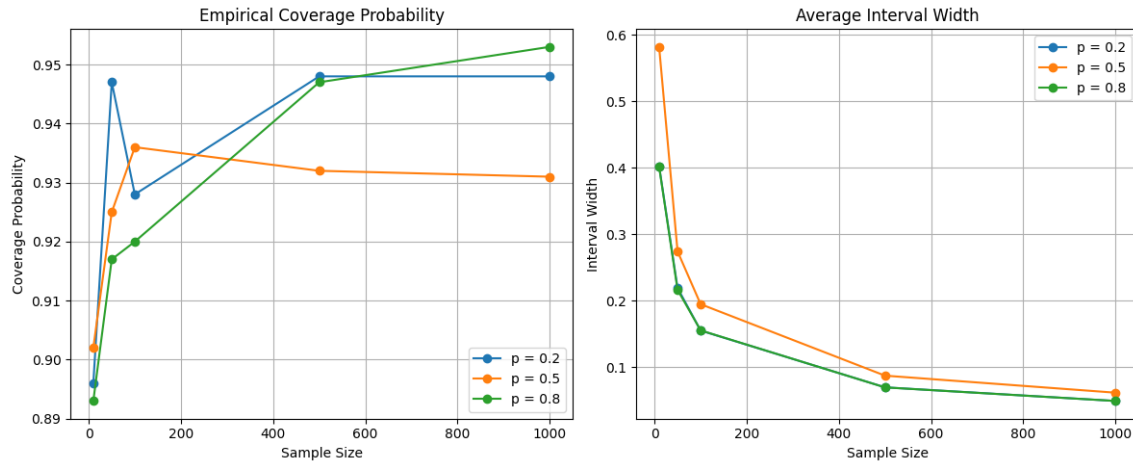| Population Proportion (p)	Sample Size	Coverage Probability	Avg. Interval Width
5*0.2	10	0.8960	0.4025
	50	0.9470	0.2193
	100	0.9280	0.1554
	500	0.9480	0.0700
	1,000	0.9480	0.0495
5*0.5	10	0.9020	0.5818
	50	0.9250	0.2743
	100	0.9360	0.1950
	500	0.9320	0.0876
	1,000	0.9310	0.0619
5*0.8	10	0.8930	0.4016
	50	0.9170	0.2166
	100	0.9200	0.1556
	500	0.9470	0.0699
	1,000	0.9530	0.0495

Figure 4: Confidence Interval Characteristics Across Different Proportions

3.3.1 Key Observations

- Coverage Probability:

    - Consistently approaches 95% theoretical level
    - Slight variations across different population proportions
    - Smallest sample size (10) shows most deviation

- Interval Width:

    - Decreases substantially with increasing sample size
    - Wider intervals for extreme proportions (0.2 and 0.8)
    - Narrowest intervals at large sample sizes (500-1,000)

3.4 Discussion

Comprehensive analysis reveals critical insights:

1. Sample Size Impact:

    - Larger samples provide more precise estimates
    - Interval width reduces significantly with increased samples
    - Convergence towards true population proportion

11

2. Proportion Sensitivity:

   - Different population proportions show varying interval characteristics
   - Extreme proportions (0.2, 0.8) display more variability
   - Midpoint proportion (0.5) demonstrates most stable estimation

3. Estimation Reliability:

   - 95% confidence level maintains consistent coverage
   - Minimal systematic bias observed across proportions

3.5 Conclusion

The study provides crucial insights into polling and proportion estimation:

- Sample size critically influences estimation precision

- Confidence intervals effectively capture true population proportions

- Larger samples reduce uncertainty in voter preference estimates

Practical Implications:

- Recommend minimum sample size of 100-500 for reliable polling

- Account for proportion variability in survey design

- Use confidence intervals to communicate estimation uncertainty