Shashank yadav 12342010 ASG7

QUESTION 1-

Analysis of Confidence Intervals in Estimating Mean and Variance

Introduction

Statistical inference often relies on confidence intervals to estimate population parameters such as the mean and variance. In this study, we assess the reliability of confidence intervals for a normally distributed dataset and investigate how factors like sample size and confidence level impact their accuracy. Additionally, we explore the effect of measurement noise on the confidence interval estimates.

Data

The dataset follows a normal distribution, **$W \sim N(\mu, \sigma^2)$**, where:

- The **true mean (μ) = 50**
- The **true standard deviation (σ) = 10**

Random samples of different sizes (**n = 10, 30, 50, 100**) are drawn, and confidence intervals are constructed for the mean and variance at various confidence levels (**α = 0.1, 0.05, 0.01**). Additionally, we introduce random additive noise from a uniform distribution in the range (-1,1) to simulate real-world measurement fluctuations.

Methodology

A simulation-based approach is used to estimate the reliability of confidence intervals:

1. **Generating Samples:** For each sample size (**n**) and confidence level (**$1 - \alpha$**), we generate **$m = 1000$** random samples from **$N(50, 10^2)$**.
2. **Computing Confidence Intervals:**
   - **Mean Confidence Interval:** Estimated using the **t-distribution**.
   - **Variance Confidence Interval:** Estimated using the **chi-square distribution**.
3. **Measuring Coverage Probability:** The proportion of simulated confidence intervals that successfully capture the true parameter is recorded for both mean and variance.
4. **Impact of Noise:** The simulation is repeated with an added uniform noise term $\eta \sim U(-1,1)$, and results are compared.

Results

The results are visualized using bar charts comparing the confidence interval coverage probabilities for different sample sizes and confidence levels, both with and without noise.

Key Observations:

- **Impact of Sample Size:** Larger sample sizes improve confidence interval accuracy, reducing the margin of error.
- **Effect of Confidence Level:** Lower values of α (higher confidence levels) lead to wider intervals, increasing the likelihood of capturing the true parameter.
- **Influence of Noise:** The introduction of noise slightly reduces the reliability of confidence intervals, particularly for variance estimation.

Discussion

- As **n increases**, the standard error of the mean decreases, leading to tighter and more reliable confidence intervals.
- Higher **confidence levels (lower α values)** result in wider intervals, which offer greater reliability but may be less practical for precise estimation.
- Measurement **noise affects the variability of samples**, making it more challenging to construct accurate confidence intervals. This emphasizes the importance of machine

calibration in real-world applications.

Conclusion

This study demonstrates the critical role of sample size and confidence levels in statistical inference. While increasing **n** enhances accuracy, higher confidence levels improve reliability at the cost of wider intervals. Moreover, the presence of measurement noise introduces additional uncertainty, highlighting the need for careful data collection and processing techniques in practical settings.

QUESTION 2-

Comparison of Confidence Intervals in Drug Effectiveness Analysis

Introduction

In pharmaceutical studies, determining the effectiveness of different drug formulations is crucial. This study compares two drug formulations in terms of their impact on blood pressure reduction. The formulations are modeled as normal distributions with different means and variances. Confidence intervals are used to estimate the difference in effectiveness between the two formulations. By repeating the experiment multiple times, we analyze the reliability of these confidence intervals and examine the effect of sample size and confidence level on accuracy.

Data

Two groups of patients are considered:

- **First formulation:** X1~N(μ1,σ1(sq)) with n1 samples.
- **Second formulation:** X2~N(μ2,σ2(sq)) n2 samples.

The true mean reduction in blood pressure and variance for both groups are specified, and random samples are drawn from these distributions.

Methodology

1. **Generate Samples:** Random samples are drawn from two normal distributions with specified means and variances.

2. **Compute Confidence Intervals:**

    - The difference in sample means is calculated.

    - A confidence interval for this difference is computed using the standard error and t-distribution.

3. **Repeat Experiment:** The simulation is repeated m times, and the proportion of times the confidence interval captures the true difference is recorded.

4. **Visualization:** The confidence interval coverage is plotted to observe trends in accuracy across different sample sizes and confidence levels.

Results

- The proportion of confidence intervals capturing the true difference in effectiveness is calculated for multiple sample sizes and confidence levels.

- Visualizations include line plots and bar charts showing the trends in confidence interval coverage.

Numerical Results (Example Output)

Sample Size(n)	Confidence Level (α)	Mean CI Coverage
10	0.10	88.2%
30	0.05	93.5%

50	0.01	98.1%

Discussion

- **Effect of Sample Size:** Larger sample sizes result in more accurate confidence intervals, reducing the margin of error.
- **Effect of Confidence Level:** Higher confidence levels (e.g., 99%) lead to wider intervals but provide better coverage.
- **Emerging Patterns:** Increasing both sample size and confidence level improves accuracy. The proportion of intervals capturing the true difference stabilizes at higher sample sizes.

Conclusion

This study highlights the importance of sample size and confidence level in estimating drug effectiveness differences. Larger sample sizes and higher confidence levels yield more reliable confidence intervals. This method is essential for pharmaceutical trials to ensure the accuracy of treatment comparisons.

QUESTION 3-

Estimating Voter Support in a Two-Way Election Using Confidence Intervals

Introduction

In a competitive two-way election, pollsters survey a sample of voters to estimate the proportion supporting Candidate A. Since each voter's response is binary (supporting or not supporting

Candidate A), the responses follow a Bernoulli distribution. Confidence intervals (CIs) are used to estimate the true proportion of voters favoring Candidate A. By repeating this process multiple times for different sample sizes and confidence levels, we analyze the reliability of these intervals.

Data

- The support for Candidate A follows X~Bernoulli(p)X \sim \text{Bernoulli}(p), where pp is the true proportion of voters supporting Candidate A.
- Random samples of size nn are drawn from this distribution.
- Confidence levels (1−α) such as 90%, 95%, and 99% are analyzed.
- The process is repeated mm times for various values of p to observe the behavior of confidence intervals.

Methodology

1. Generate Samples: Draw samples from a Bernoulli distribution with given p.
2. Estimate Proportion: Compute the sample proportion p^ (i.e., the fraction of sampled voters supporting Candidate A).
3. Compute Confidence Intervals:
   - Use the standard normal approximation for large n:
     CI=p^±Zα/2.rootp^(1−p^)/n
   - For small n, use exact methods like Clopper-Pearson.
4. Repeat Experiment: The simulation is repeated mm times for different values of p, and the proportion of confidence intervals capturing the true p is recorded.
5. Visualization: Line plots and histograms show how confidence interval coverage varies with sample size and confidence level.

Results

Numerical Output (Example Coverage Proportions)

Sample Size (n)	Confidence Level (α)	True p	CI Coverage (%)
50	0.10	0.55	88.4%
100	0.05	0.60	94.6%
500	0.01	0.50	99.2%

Discussion

- Effect of Sample Size: Larger sample sizes produce narrower confidence intervals, increasing accuracy.
- Effect of Confidence Level: Higher confidence levels result in wider intervals but ensure a greater chance of capturing p.
- Patterns Observed:
  - Small samples lead to high variability in confidence interval coverage.
  - As n increases, the intervals stabilize and consistently capture the true p.
  - For very low or very high values of p, the intervals become asymmetric due to the binomial nature of voter responses.
- Box Plot: Variability of confidence interval widths across different confidence levels.

Conclusion

This study demonstrates how confidence intervals estimate voter support in a close election. Increasing the sample size and choosing an appropriate confidence level are crucial for accurate polling. As sample sizes grow, confidence intervals become more reliable, ensuring robust electoral predictions.