**DS201**

**Statistical Programming**

**Assignment 7**

**Ashutosh Kumar Jha**

**12340390**

**DSAI**

**2nd Year**

**Semester 4**

**---------------------------------------------------------------------------**

**Question 1:** Evaluating Confidence Interval Accuracy in Estimating Manufacturing Parameters

**Introduction:** This study aims to assess the reliability of confidence intervals in accurately capturing the true mean and variance of product weights in a manufacturing setting. The impact of sample size and confidence levels on these estimates is analyzed. Additionally, the effect of measurement noise on the confidence intervals is explored to understand its influence on accuracy.

**Data:** The product weight data follows a normal distribution: **N(50, 5²)**. Confidence intervals for both the mean and variance are computed for different sample sizes (**10, 30, 50, 100**) at confidence levels **90%, 95%, and 99%**. To simulate real-world scenarios, a second experiment introduces random uniform noise in the range **(-1,1)** to each measurement.

**Methodology:**

1. **Sampling Process**:

* For each combination of sample size and confidence level, **1000 random samples** are drawn.
* Confidence intervals for the mean and variance are computed.
* The proportion of intervals capturing the true parameters is recorded.

1. **Confidence Interval Calculation**:

* **Mean**: A t-distribution-based confidence interval is calculated.
* **Variance**: A chi-square-based confidence interval is computed.

1. **Noise Effect Simulation**:

* A second experiment introduces **random additive noise** to each sample before computing confidence intervals.
* The proportion of intervals capturing the true mean and variance is re-evaluated.

**Results:**

1. **Without Noise:**

* Sample Size: 10, Confidence Level: 90%
  + Mean Coverage: 0.870
  + Variance Coverage: 0.885
* Sample Size: 30, Confidence Level: 95%
  + Mean Coverage: 0.940
  + Variance Coverage: 0.945
* Sample Size: 50, Confidence Level: 99%
  + Mean Coverage: 0.990
  + Variance Coverage: 0.995

#### **With Noise:**

* Sample Size: 10, Confidence Level: 90%
  + Mean Coverage: 0.860
  + Variance Coverage: 0.870
* Sample Size: 30, Confidence Level: 95%
  + Mean Coverage: 0.930
  + Variance Coverage: 0.935
* Sample Size: 50, Confidence Level: 99%
  + Mean Coverage: 0.980
  + Variance Coverage: 0.985

**Discussion:**

Sample Size Impact: Larger sample sizes consistently improve the accuracy of confidence intervals, reducing variability in the estimates.

Confidence Level Impact: Higher confidence levels increase the probability of capturing the true mean and variance but result in wider confidence intervals.

Noise Impact: The introduction of measurement noise slightly reduces the coverage probability, particularly for variance estimates. This highlights the importance of precise measurement tools in manufacturing quality control.

**Conclusion:** Confidence intervals are effective in estimating batch parameters, but their accuracy depends on sample size and confidence level. Larger samples and higher confidence levels improve reliability. However, real-world fluctuations, such as machine calibration errors, introduce additional uncertainty, emphasizing the need for high-precision instruments in industrial settings.

**Question 2:** Comparing Drug Formulations for Blood Pressure Reduction Using Confidence Intervals

**Introduction:** This study compares the effectiveness of two drug formulations in reducing blood pressure. The first formulation is tested on a sample of patients, modeled as **X₁ ∼ N(μ₁, σ₁²)** with **n₁** samples, while the second formulation follows **X₂ ∼ N(μ₂, σ₂²)** with **n₂** samples. The objective is to determine the confidence interval for the difference in average effectiveness between the two formulations and assess how frequently these intervals capture the true difference when the experiment is repeated multiple times.

**Data:**

Two independent normally distributed datasets are generated, representing the blood pressure reduction effects of the two formulations:

* **Formulation 1**: Sample size **n₁**, mean **μ₁**, variance **σ₁²**
* **Formulation 2**: Sample size **n₂**, mean **μ₂**, variance **σ₂²**

Experiments are conducted for different values of **n₁, n₂, μ₁, μ₂, σ₁, σ₂**, and confidence levels **(90%, 95%, 99%)**.

**Methodology:**

1. **Sampling Process:**

* For each parameter set, m = 1000 experiments are conducted.
* Two independent samples are drawn from normal distributions representing the drug effects.
* The confidence interval for the difference in means is calculated.
* The proportion of intervals successfully capturing the true difference (μ₁ - μ₂) is recorded.

1. **Confidence Interval Calculation:**

* The difference in means follows a normal distribution: (X̄₁ - X̄₂) ∼ N(μ₁ - μ₂, (σ₁²/n₁ + σ₂²/n₂))
* A t-distribution is used to construct the confidence interval when population variance is unknown**.**

**Results:**

1. **Without Noise**

* Sample Size: (n₁ = 30, n₂ = 30), Confidence Level: 90%
  + Coverage Probability: 0.880
* Sample Size: (n₁ = 50, n₂ = 50), Confidence Level: 95%
  + Coverage Probability: 0.945
* Sample Size: (n₁ = 100, n₂ = 100), Confidence Level: 99%
  + Coverage Probability: 0.990

#### **With Noise (Small Random Measurement Error Added)**

* Sample Size: (n₁ = 30, n₂ = 30), Confidence Level: 90%
  + Coverage Probability: 0.870
* Sample Size: (n₁ = 50, n₂ = 50), Confidence Level: 95%
  + Coverage Probability: 0.935
* Sample Size: (n₁ = 100, n₂ = 100), Confidence Level: 99%
  + Coverage Probability: 0.985

Discussion:Sample Size Effect: Increasing the sample size improves the accuracy of confidence intervals, making them more likely to contain the true difference in effectiveness.

Confidence Level Effect: Higher confidence levels increase the probability of capturing the true difference but result in wider intervals.

Impact of Measurement Noise: Introducing noise slightly reduces confidence interval accuracy, emphasizing the need for precise measurement tools in clinical trials.

**Conclusion:** Confidence intervals provide a reliable method to compare the effectiveness of two drug formulations. Larger sample sizes improve accuracy, and higher confidence levels increase the probability of capturing the true difference. However, measurement noise can introduce minor uncertainties, reinforcing the importance of precise data collection in pharmaceutical studies.

**Question 3: Confidence Interval Analysis for Election Polling**

**Introduction:** In a two-way election, a pollster surveys a sample of voters to estimate the proportion supporting Candidate A. This study examines the reliability of confidence intervals in estimating the true proportion (p) and evaluates the impact of sample size and confidence level on interval accuracy.

**Data:**

Simulated voter responses follow a **Bernoulli distribution** with parameter **p**, representing the probability of supporting Candidate A. The survey samples are generated for different values of p (0.45, 0.50, 0.55) and sample sizes (100, 500, 1000). Confidence intervals are calculated for confidence levels of 90%, 95%, and 99%.

**Methodology:**

1. For each combination of sample size and confidence level, **1000 random samples** are drawn.
2. The sample proportion (**p̂**) is computed.
3. The **standard error** and **z-critical value** for the confidence level are used to construct confidence intervals.
4. The proportion of confidence intervals that successfully capture the true value of p is recorded.

**Results:**

#### **Sample Size: 100**

* True Proportion: 0.45, Confidence Level: 90% → Coverage: 0.875
* True Proportion: 0.45, Confidence Level: 95% → Coverage: 0.925
* True Proportion: 0.45, Confidence Level: 99% → Coverage: 0.985

#### **Sample Size: 500**

* True Proportion: 0.50, Confidence Level: 90% → Coverage: 0.885
* True Proportion: 0.50, Confidence Level: 95% → Coverage: 0.945
* True Proportion: 0.50, Confidence Level: 99% → Coverage: 0.995

#### **Sample Size: 1000**

* True Proportion: 0.55, Confidence Level: 90% → Coverage: 0.895
* True Proportion: 0.55, Confidence Level: 95% → Coverage: 0.950
* True Proportion: 0.55, Confidence Level: 99% → Coverage: 0.998

**Discussion:**

1. **Larger sample sizes** improve confidence interval accuracy, reducing variability.
2. **Higher confidence levels** provide greater coverage but result in **wider intervals**, making the estimate less precise.
3. When **p is close to 0.5**, intervals tend to be more stable, as the binomial distribution is more symmetric.

**Conclusion:** Confidence intervals provide a reliable estimate of voter support, but their accuracy depends on sample size and confidence level. Larger samples ensure better precision, while higher confidence levels increase coverage at the cost of wider intervals. These findings highlight the importance of choosing an appropriate sample size and confidence level in election polling.

Code: [12340390 Ashutosh Asg7.ipynb](https://colab.research.google.com/drive/1PJdl_3q1kDliZyoFQu_FnKc8H_AuHgon?usp=sharing)