Title: The Power of Sampling in Statistical Analysis

Subtitle: Leveraging Statistical Sampling to Assess Air Quality

ISSUE / PROBLEM

This project aimed to analyze air quality data, specifically carbon monoxide levels, across multiple locations. Given the large dataset, a direct analysis of the entire population was not feasible. Therefore, the objective was to apply statistical sampling techniques to obtain reliable insights while maintaining efficiency and accuracy in reporting findings.

IMPACT

The findings reinforce the power of statistical sampling in environmental studies and public health policy:

- The **AQI values in the sample were statistically below the threshold of 100**, indicating satisfactory air quality in the examined locations.
- This statistical approach allows for informed decision-making while reducing the resources needed for full-population analysis.
- Future efforts should focus on regions where AQI may exceed safe levels, optimizing funding allocation for environmental interventions.

RESPONSE

This project utilized sampling techniques, including sampling with replacement, to extract representative subsets from the dataset. The analysis followed a structured approach:

- A sample of **50 AQI values** was drawn to estimate population-level statistics.
- The **mean AQI from the dataset was 6.76**, and the **mean of the sample was 5.54**, showcasing expected sampling variability.
- The Central Limit Theorem was applied, demonstrating that **the distribution of sample means approximates a normal distribution** when repeated sampling is conducted.
- A **standard error of 0.7413** was calculated, confirming statistical reliability in estimation.

KEY INSIGHTS

- **Sampling provides a valid representation of population trends** without analyzing every data point.
- **The mean of the sample closely aligns with the population mean**, validating the efficacy of the sampling approach.
- **The sampling distribution follows a normal distribution**, as predicted by the Central Limit Theorem.
- **Statistical significance is established**, as the sampled AQI values were at least **2–3 standard errors below the critical threshold** of 100.