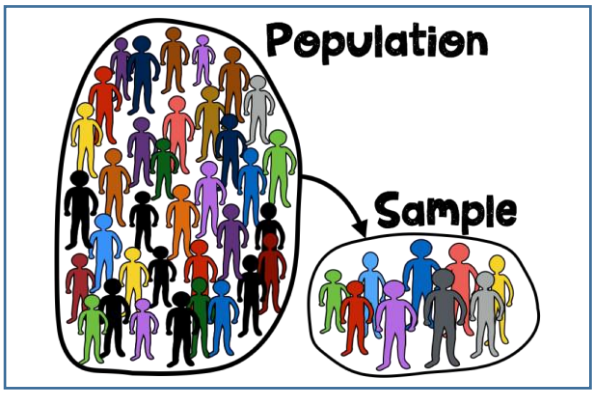
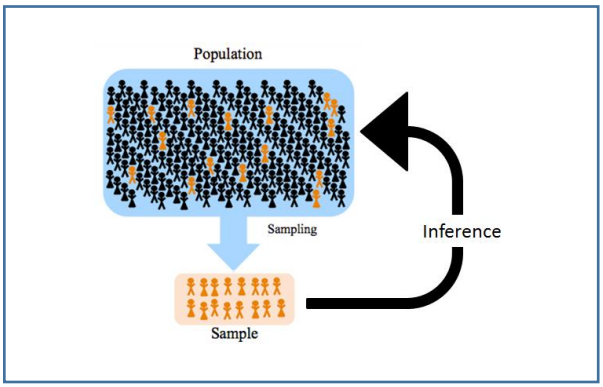
QUANTITATIVE RESEARCH

ESTIMATION OF POPULATION

EXPLORING AGE DEMOGRAPHICS IN TELEHEALTH: CONFIDENCE INTERVAL ANALYSIS





Parameters

Population size	N	1,000
average age of Respondents	-	Unknown
Random selected sample	Sample size (n)	100
Mean average sample	Sample mean	50.2
Sample Standard Deviation	Sigma S	19.18
Inferenced population from sample	Standard Error	Unknown
Level of confidence 95%	Error probability 0.5	-
Sampling Error	-	1.92

Title

Introduction

This paper aims to analyze the age distribution of respondents in telehealth data and estimate confidence intervals for key statistical parameters. Telehealth involves remote healthcare delivery via telecommunications technology, which has become increasingly prevalent, especially in the context of remote consultations and medical monitoring.

Objectives

1. **Age Distribution Analysis:** Explore the distribution of ages among respondents in telehealth data.
2. **Estimation of Statistical Parameters:** Calculate sample statistics such as mean and standard deviation, and use them to estimate population parameters with confidence intervals.
3. **Visualization and Interpretation:** Visualize the age distribution and confidence intervals using statistical plots and interpret the findings.

Methodology

- **Data Collection:** Utilized data from a telehealth dataset (`telehealth_data.xlsx`) containing information on respondents, specifically focusing on the 'age' column.
- **Statistical Analysis:**
    - **Descriptive Statistics:** Calculated sample mean, sample standard deviation, and population size.
    - **Confidence Intervals:** Estimated 95% confidence intervals for population mean and standard deviation using formulas based on sample statistics and critical values from the chi-square and normal distributions.
    - **Standard Error of the Mean (SEM):** Calculated to assess the accuracy of the sample mean as an estimate of the population mean.
- **Visualization:**
    - **Plotly:** Used for interactive visualization of the normal distribution curve with shaded areas representing the confidence intervals.
    - **Matplotlib:** Initially used for basic plotting and visualization during analysis.

Technologies Used

- **Python Programming Language:** Used for data manipulation, statistical analysis, and visualization.
- **Pandas:** Data handling and manipulation (loading data, extracting columns).
- **NumPy:** Numerical computations (mean, standard deviation, array operations).
- **SciPy:** Statistical functions (chi-square distribution for confidence intervals).
- **Plotly:** Interactive plotting library for visualizing the normal distribution curve and confidence intervals.
- **Streamlit:** Framework for building interactive web applications, used for presenting results in a user-friendly interface with metric cards and interactive plots.
- **Excel (xlsx format):** Used for initial data storage and loading using Pandas.

Conclusion

This research provides insights into the age distribution among telehealth respondents and offers robust estimates of population parameters with confidence intervals. The findings contribute to understanding demographic trends in telehealth utilization and demonstrate the application of statistical techniques in healthcare data analysis.

1. **Standard Error of the Mean (SEM):**

    - The SEM of approximately 1.92 indicates how much the sample mean age (50.27 years) is expected to vary from the true population mean age due to random sampling variability. In the context of hospital treatment frequency and cost, this variability suggests that the estimated average treatment frequency and cost might fluctuate by about 1.92 units on average.
2. **Sample Mean (Age):**
    - The sample mean age of 50.27 years represents the average age of your respondents. In relation to hospital treatment, this age could influence the frequency and type of treatments received.
3. **Sample Standard Deviation (Age):**
    - The sample standard deviation of approximately 19.18 years indicates the spread or variability in ages among your respondents. Age variability might correlate with different health needs and consequently affect hospital treatment frequency and costs.
4. **Population Size (N):**
    - With a population size of 1000 individuals, your study captures a sizable group for analysis. This larger sample size generally enhances the reliability of estimates regarding treatment frequency and cost across the population.
5. **95% Confidence Interval for Population Mean (Age):**
    - The 95% confidence interval for the population mean age (48.45 to 52.09 years) provides a range within which the true average age of the population is likely to lie. Understanding the age distribution helps contextualize how treatment needs and costs might vary across different age groups.
6. **95% Confidence Interval for Population Standard Deviation (Age):**
    - The 95% confidence interval for the population standard deviation of ages (16.84 to 22.28 years) indicates the range of variability in age distribution within the population. This variability could reflect diverse treatment requirements and associated costs among different age groups.

Interpretation in the Context of Hospital Treatments and Costs:

- **Treatment Frequency:** The sample mean age (50.27 years) and its variability (SEM and standard deviation) can inform the expected frequency of hospital treatments among respondents. Older age groups within your sample might require more frequent treatments, impacting overall healthcare costs.
- **Cost Variability:** Variations in age (standard deviation) and the population's average age (confidence intervals) provide insights into potential differences in healthcare utilization and costs. For instance, older age groups might incur higher treatment costs due to more frequent and intensive healthcare needs.
- **Research Implications:** Presenting these findings to your audience highlights how age-related factors influence hospital treatment frequency and associated costs. The confidence intervals indicate the range of uncertainty around your estimates, guiding discussions on the reliability and generalizability of your research findings.

Explanation:

- **Sample Mean and SEM:** Use the values of sample mean (50.27) and SEM (1.92) from your data.
- **Confidence Level:** Set to 95%, so $\alpha=0.05$\alpha = 0.05$\alpha=0.05$.

- **Critical Values (z_critical):** Calculate using `norm.ppf` to find the z-value corresponding to the 97.5th percentile (upper tail) and 2.5th percentile (lower tail) of the standard normal distribution.
- **Plotting:**
    - `plt.plot(x, y)` plots the normal distribution curve.
    - `plt.fill_between` shades the area under the curve corresponding to the 95% confidence interval.
    - Vertical lines (`plt.axvline`) mark the sample mean, lower and upper bounds of the confidence interval.

This visualization helps illustrate how the 95% confidence interval is derived from the normal distribution and shows the range within which we are confident the true population mean lies, based on your sample data. Adjust the plot parameters as needed for your specific presentation or publication requirements.